Graduate Texts in Mathematics

David Eisenbud

Commutative Algebra with a View Toward Algebraic Geometry

Springer-Verlag

Graduate Texts in Mathematics 150

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces.
4 HILTON/STAMMBACH. A Course in Homological Algebra.
5 MAC LANE. Categories for the Working Mathematician.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 WERMER. Banach Algebras and Several Complex Variables. 2nd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I. 2nd ed.

continued after index

David Eisenbud

Commutative Algebra

with a View Toward Algebraic Geometry

With 90 Illustrations

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest

David Eisenbud
Department of Mathematics
Brandeis University
Waltham, MA 02254
USA

Mathematics Subject Classifications (1991): 13-01, 14-01, 13A50, 13C15

Library of Congress Cataloging-in-Publication Data
Eisenbud, David.
Commutative algebra with a view toward algebraic geometry/David Eisenbud.
p. cm. – (Graduate texts in mathematics; v. 150)
Includes bibliographical references and index.
ISBN-13: 978-3-540-78122-6 e-ISBN-13: 978-1-4612-5350-1
DOI: 10.1007/ 978-1-4612-5350-1

1. Commutative algebra. 2. Geometry, Algebraic. I. Title.
II. Series.
QA251.3.E38 1994
512′.24–dc20 94-17351

Printed on acid-free paper.

Softcover reprint of the hardcover 1st edition 1995

Production managed by Natalie Johnson; manufacturing supervised by Jacqui Ashri.
Photocomposed using LaTeX files supplied by Laser Words, Madras, India.

9 8 7 6 5 4 3 2 1

To Leonard, Ruth-Jean, Monika:
For capacity, passion, endurance.

Contents

[1]The selected exercises are marked with a *.

Introduction

> ... I was not able to write anything about it [bullfighting] for five years—and I wish I would have waited ten. However, if I had waited long enough I probably never would have written anything at all since there is a tendency when you really begin to learn something about a thing not to want to write about it but rather to keep on learning about it always and at no time, unless you are very egotistical, which, of course, accounts for many books, will you be able to say: now I know all about this and will write about it. Certainly I do not say that now; every year I know there is more to learn ... .
>
> —Ernest Hemingway, from "Death in the Afternoon."[1]

It has seemed to me for a long time that commutative algebra is best practiced with knowledge of the geometric ideas that played a great role in its formation: in short, with a view toward algebraic geometry.

Most texts on commutative algebra adhere to the tradition that says a subject should be purified until it references nothing outside itself. There are good reasons for cultivating this style; it leads to generality, elegance, and brevity, three cardinal virtues. But it seems to me unnecessary and undesirable to banish, on these grounds, the motivating and fructifying ideas on which the discipline is based.

---

[1]Reprinted with permission of Scribner, an imprint of Simon & Schuster, from *Death in the Afternoon* by Ernest Hemingway. Copyright 1932 by Charles Scribner's Sons. Copyright renewed © 1960 by Ernest Hemingway.

In this book I have tried to write on commutative algebra in a way that makes the heritage of the subject apparent. I have allowed myself many words and pictures with the vague and difficult aim of clarifying the "true meaning" of the results and definitions. For all this, I have tried not to compromise the technical perfection to which the subject has been brought by masters like Hilbert, Emmy Noether, Krull, Van der Waerden, and Zariski, to name only a few of those no longer living.

Advice for the Beginner

Because of my attempt to mix algebra and geometry, this text has a certain unevenness of level. Dear reader, unless you are unusually experienced, you will probably find some passages for which you are simply unprepared, a problem you would not encounter with a book written in a more linear style. You should feel free to skip lightly over, or "read for culture," explanatory material which seems difficult, or which uses ideas of which you have not yet heard. Perhaps when you do hear of them—and you will, as they come from the mainstream—you will feel a sense of recognition, knowing that they have something to do with this subject. I have taken some pains to make a thread of theorems and definitions that are stated without reference to these more obscure passages. You should think of them as something to return to when more of the pieces in the vast puzzle of mathematics have fallen into place for you.

Information for the Expert

I shall now describe some of the contents of this book, emphasizing its more novel features. From the beginning, my goal has been to cover at least the material that graduate students studying algebraic geometry—and in particular those studying *Algebraic Geometry*, the excellent book by Robin Hartshorne [1977]—should know (in fact the title of this book began as a pun). In particular, all the algebraic results referred to in that book without proof may be found here.

The first chapter sets the stage: It surveys some of the prehistory of commutative algebra in number theory, the theory of Riemann surfaces, and invariant theory; and it concludes with a survey of Hilbert's amazing contributions near the end of the nineteenth century. I have done this to provide something interesting right at the beginning and to introduce the reader to the translation between commutative algebra and the geometry of affine and projective varieties. Much use is made of this translation later in the book, though mostly in a very elementary way. Chapter 1 also introduces graded rings, to which we return often.

The second chapter begins afresh, with that now indispensable operation, localization. The chapter includes an analysis of rings whose primes are all maximal—what are later called zero-dimensional rings.

Chapter 3 on primary decomposition begins with the standard treatment, emphasizing associated primes. Symbolic powers and their connections with the order of vanishing of functions (the theorem of Nagata and Zariski) are discussed to provide a nontrivial application. I also discuss the geometric information hidden in the embedded components. The exercises include a complete treatment of primary decomposition for monomial ideals, a number of examples, and an exploration of the nonuniqueness of embedded components.

Chapter 4 concerns the Nullstellensatz and integrality. I develop Nakayama's lemma here from the Cayley-Hamilton theorem, and study the behavior of primes in an integral extension—the relative version of the zero-dimensional theory treated in Chapter 2. Five different proofs of the Nullstellensatz are given in this book: The text of Chapter 4 contains the strongest, which is essentially due to Bourbaki. The exercises treat the proof by Artin-Tate and two "quick-and-dirty" methods, one due to Van der Waerden and Krull and one for which I don't know an attribution; I learned it from Artin. The fifth proof, using the Noether normalization theorem, is given in Chapter 13.

Chapter 5 takes up some of the constructions of graded rings from a ring and an ideal: the associated graded ring and the "blowup algebra." The Krull intersection theorem is proved there.

Chapter 6 is concerned with flatness. A number of simple geometric examples are intended to convey the notion that flatness is a kind of "continuity of fibers." I then take up a number of characterizations of flatness, for example the one by equations, and the "local criterion." This chapter also contains a gentle introduction to the use of Tor.

I next treat the concept of completion, emphasizing the good geometric properties that come from Hensel's lemma. I present completion as a sort of superlocalization that allows one to get at neighborhoods much smaller than a Zariski neighborhood. Hensel's lemma is presented as a version of Newton's method for finding solutions to equations. There is a thorough treatment of coefficient fields and the equicharacteristic part of the Cohen structure theorems.

Chapter 8 begins the treatment of dimension theory. I begin with a survey, to explain some history and bring forward the main points of the theory. I even give a set of axioms characterizing Krull dimension, hoping in this way to explain the central role of the theorems about the dimension of fibers. This chapter is somewhat more advanced than the ones around it and is meant to be read "for culture only" on a first pass through the subject. Nothing in it is required for the subsequent development.

In the following chapter, therefore, I have repeated some of the most basic definitions and also collected the information about dimension that

was accumulated (without an appropriate language) in earlier parts of the book —essentially the theory of dimension zero and relative dimension zero.

Chapter 10 handles the principal ideal theorem (I give Krull's proof) and its consequences. This is where regular local rings and regular sequences are introduced. The fact that a regular local ring is a domain is proved as an application. The exercises contain, among other things, a treatment of the codimensions of determinantal ideals.

Chapter 11 treats "dimension and codimension one"—that is, essentially, normal rings (including discrete valuation rings and Serre's criterion) and the ideal class group. Dedekind domains are treated along the way.

Chapter 12 introduces the Hilbert-Samuel function and polynomial; the easy case of the Hilbert function and polynomial was already presented in Chapter 1. Multiplicities naturally appear here.

Chapters 13 and 14 take up a somewhat deeper side of dimension theory, examining affine rings and the dimensions of fibers of finitely generated algebras. I explain something of classical as well as modern elimination theory.

In Chapter 15 I give an account of the theory of initial ideals and Gröbner bases, including the theorems of Galligo, Bayer and Stillman on generic initial ideals. Relative to the other presentations available I take a rather mathematical approach to the subject. I feel that this leads to considerable simplification without sacrificing the power to "actually compute" that this theory affords. At the end of the chapter is a long series of applications and a set of computer algebra "projects" showing how the computational possibilities of this theory let one make new conjectures, hard and easy, trivial as well as significant.

Chapter 16 is about modules of differentials. My goals are to explain the roles these play in linearizing problems, from the Jacobian criterion to infinitesimal automorphisms to deformation theory, and also to prove some of the technical results that intervene in the field theory necessary for the Cohen structure theorems and for various topics concerning finitely generated algebras (separability, p-bases, differential bases).

The final chapters treat and use the homological tools in earnest. I begin with an elementary treatment of the Koszul complex of two elements. (This is adapted from the treatment by David Buchsbaum that first lured me into commutative algebra 25 years ago.) Next follows a technical account of the Koszul complex, using some multilinear algebra. In the exercises, among other things, are Priddy's generalized Koszul complex (an explicit form for the linear part of the resolution of the residue class field) and the Taylor complex (a resolution of monomial ideals).

The notion of depth and the Cohen-Macaulay property occupy Chapter 18. After establishing the basic properties, such as localization, I explain applications of the Cohen-Macaulay property: Macaulay's unmixedness theorem; Hartshorne's theorem on connectedness in codimension one; flatness over a regular base; and the application to proving that

an ideal is prime, using Serre's characterization of normality.

The homological characterization of regular local rings as those of finite global dimension is presented in Chapter 19, along with the application to factoriality. This requires some talk of stable freeness, and I present the classic example of the tangent bundle to the real 2-sphere. The Auslander-Buchsbaum formula and the associated characterization of Cohen-Macaulay rings are here too.

Chapter 20 examines a number of topics concerning free resolutions. Various criteria of exactness are presented. The material is approached through the Fitting invariants and their significance. I present the Hilbert-Burch theorem characterizing ideals of projective dimension 1, and apply this to finding the equation the cubic surface in $\mathbf{P}^3$ corresponding to six given points in the plane. The chapter closes with an algebraic treatment of Castelnuovo-Mumford regularity. The expert reader will recognize that the selection of material for this chapter has much to do with my personal taste and experience.

Chapter 21 contains an account of the canonical module and duality for local Cohen-Macaulay rings, and some of the theory of Gorenstein rings. I have included more than the usual amount of material on the Artinian case (including "pictures" of the canonical module), with a view to giving the student some comfort in that case and motivating the use of injective dimension in the general case. The canonical module is defined as a module that reduces, modulo a regular sequence, to the canonical module of the associated Artinian ring. This treatment seems to me somewhat more concrete and accessible than the one found in most other expositions. As an application I explain something of linkage. The exercises contain a proof of the Cayley-Bacharach theorem in a modern formulation.

Throughout the text I have tried to include illustrations of the power of the ideas on concrete examples provided by geometry. For example, I illustrate the Hilbert-Burch theorem not only with the application to cubic surfaces, but also, in Chapter 21 for the proof by Apéry and Gaeta that Cohen-Macaulay ideals of codimension two in a regular ring are linked to complete intersections.

It is hard to do commutative algebra without knowing at least a small amount of field theory (separable extensions, p-bases), category theory (functors, natural transformations, adjointness, limits, and colimits), homological algebra (projective and injective resolutions, Tor, Ext, and local cohomology), and multilinear algebra (symmetric and exterior algebras). I have provided appendices on these subjects that far exceed the actual requirements for this course. For example, the appendix on limits and colimits contains a treatment of the Lazard-Govorov characterization of flat modules; and the appendix on multilinear algebra contains a treatment of the Eagon-Northcott "family" of complexes, sufficiently thorough to allow the reader to write down, for example, explicit minimal free resolutions for the ideals of elliptic normal curves.

The last appendix outlines enough local cohomology to explain the algebraic interpretation of the cohomology of coherent sheaves on projective spaces.

The exercises contain a large number of theoretical results, worked out as sequences of problems. I personally don't like hard exercises very much; why spend time on them rather than on doing research? So I have tried to break the problems down into fairly small pieces. Many basic geometric objects, such as toric varieties, are also illustrated. In general, I have used the exercises to expose some of the topics I have omitted from the text (the fact that the reader can have the fun of "inventing" these topics, with guidance, seems to me a positive effect of the inevitable lack of space). At the end of the book I have provided hints or sketches of solutions to quite a few of the exercises, indicated by a *. In those few cases where I have later used the result of an exercise, a reasonably full solution is given.

Prerequisites

The formal prerequisites for reading this book are rather modest, although because of the mixing of subjects a certain sophistication is necessary for reading it without the help of a teacher. I have presupposed a background in algebra on the level of a good undergraduate preparation: knowledge of groups, rings, fields, and abstract vector spaces. For the later sections of the part on dimension theory, a little Galois theory is required. All the necessary facts from homological algebra that are not included in the main text are developed from scratch in Appendix 3, but the reader who has never heard of Ext and Tor before may find this treatment rather compressed. It is not necessary to follow the more demanding sections on geometry in order to understand the rest of the book; but in order to enjoy them one needs to know such things as what a tangent space is and what the implicit function theorem says, and also something about analytic functions. I see the most natural reader of this book as one who has taken courses in algebra, geometry, and complex analysis at the level of a first-year graduate program. However, the actual knowledge required is much less, and it is possible to tackle most of the book with only an undergraduate preparation in algebra.

Sources

Standard references for some of the material treated here are the books of Zariski and Samuel [1958], Serre [1957], Bourbaki [1983, 1985], Atiyah and MacDonald [1969], Kunz [1985], and Matsumura [1980, 1986]. I have often leaned on the extremely elegant but resolutely nongeometric treatment of

Kaplansky [1970], from whom I first heard about many of the theorems presented here, and on the deep and beautiful book of Nagata [1962]. The books of Matsumura are perhaps the best general references for the subject, but are difficult for beginners (and weakly motivated algebraic geometers). The books of Kunz [1980] and Peskine [in press] share a geometric slant with this one, but differ from it in content and style. The book of Stückrad and Vogel [1986] contains extensive material on Buchsbaum rings and linkage not found in the other treatments mentioned, with a wealth of references to the literature. The new book of Bruns and Herzog [1993] contains an up-to-date treatment of the homological and module-theoretic aspects of commutative algebra. The undergraduate book by Reid (not yet out as of this writing) shares some of the spirit of this book, but covers much less material. The book of Cox, Little, and O'Shea [1992] does a particularly nice job of explaining, at an undergraduate level, the relation of geometry with the algebra of polynomial rings. It contains an excellent treatment of Gröbner bases, more elementary than the one presented in Chapter 15 of this book. The early chapters of Fulton's book [1969] on algebraic curves is another excellent source for the connection between algebra and geometry. I am grateful to the authors of these books, having learned from them.

For the history of the subject I have leaned heavily on the account of nineteenth-century number theory, invariant theory, and algebraic geometry given by Morris Kline [1972], and also on the historical summaries in the books of Krull [1968], Nagata [1962], Bourbaki [1983, 1985], and Edwards [1977]. Some material on topological dimension theory comes from Hurewicz and Wallman [1941].

Courses

There are at least two natural one-semester courses that can be made from this book, corresponding roughly to the first and second halves. Here are possible syllabi. The assignments are a plausible (though not canonical) minimum; I would expect any instructor to add, according to taste, and I would probably make a different minimum set myself each time I taught the book.

A First Course

For students with no previous background in commutative algebra, this course covers the basics through completions, some of Cohen structure theory, and a thorough treatment of dimension theory.

Chapter 1: Roots of Commutative Algebra. Do 1.2–1.4 and 1.11; more depending on the experience of the students. (Assign the rest as reading.) Exercises: 1.1–1.4, 1.18, 1.19, 1.22, 1.23

Chapter 2: Localization. All but "Products of Domains."
Exercises: 2.3, 2.4, 2.6, 2.11, 2.15, 2.19, 2.26

Chapter 3: Associated Primes and Primary Decomposition. All but "Symbolic Powers ..." and "A Determinantal Example."
Exercises: 3.1, 3.3, 3.4, 3.6

Chapter 4: Integral Dependence and the Nullstellensatz. All.
Exercises: 4.1, 4.3, 4.4, 4.9, 4.13, 4.20, 4.24, 4.29

Chapter 5: Filtrations and the Artin-Rees Lemma. All.
Exercises: 5.3, 5.5

Chapter 6: Flat Families. Through Corollary 6.3.
Exercises: 6.1, 6.4, 6.7, 6.9, 6.12

Chapter 7: Completions and Hensel's Lemma. 7.1–7.6. Concentrate on Hensel's lemma. Do statement of Cohen structure theorem (7.7); do coefficient fields only in characteristic 0; skip the proof of the Cohen structure theorem.
Exercises: 7.1, 7.5, 7.6, 7.8, 7.9, 7.19, 7.20, 7.25

Chapter 8: Introduction to Dimension Theory. As much as will fit in one lecture, stressing fibers (Axiom D3) and Theorems A, B, and C.

Chapter 9: Fundamental Definitions of Dimension Theory. All.
Exercises: 9.1, 9.2 (prepares for the proof of Noether normalization), 9.3, 9.4

Chapter 10: The Principal Ideal Theorem and Systems of Parameters. All.
Exercises: 10.1, 10.4, 10.5, 10.9, 10.10

Chapter 11: Dimension and Codimension one. Through 11.6. Sections on invertible modules, class group, Dedekind domains as time permits. Skip section on multiplicity of principal ideals.
Exercises: 11.1, 11.7, 11.8, 11.10, 11.13

Chapter 12: Dimension and Hilbert-Samuel Polynomials. All.
Exercises: 12.1, 12.2, 12.5

Chapter 13: The Dimension of Affine Rings. All.
Exercises: 13.1, 13.2, 13.3, 13.6, 13.12, 13.13

Chapter 14: Elimination Theory, Generic Freeness, and the Dimension of Fibers. As time permits.
Exercises: 14.1, 14.5, 14.8

If time permits one further topic, my choice would be Chapter 15: Gröbner Bases, through Algorithm 15.9, as this allows the computation of dimension for affine (especially graded) rings. This chapter can also serve as the text of a short course in computational commutative algebra. For exercises, see below.

A Second Course

For students whose preparation includes something like the contents of Atiyah and MacDonald [1969] or a course like the first course just described and a small amount of homological algebra, here is a course covering

Gröbner basis techniques of computation, homological methods, some theory of free resolutions, Gorenstein rings, and duality. Differentials and the Jacobian criterion would be an option.

Review of multilinear algebra, as required: (Sections A2.1–A2.3).
Exercises: A2.2, A2.7

Review of free resolutions, Ext and Tor, as required: (Sections A3.9–A3.11).
Exercises: A3.16, A3.17, A3.18, A3.23, A3.26

Chapter 15: Gröbner Bases. Through Corollary 15.11 (proof of the Hilbert-Syzygy theorem).
Exercises 15.3, 15.4, 15.5, 15.14, 15.27, 15.29, 15.30

Option: Chapter 16: Modules of Differentials. Through Theorem 16.19 (Jacobian criterion).
Exercises: 16.1, 16.2, 16.3, 16.7, 16.8

Chapter 17: Regular Sequences and the Koszul Complex. Through Proposition 17.14.
Exercises: 17.2, 17.7, 17.12, 17.15, 17.16

Chapter 18: Depth, Codimension, and Cohen-Macaulay Rings. Through 18.15.
Exercises: 18.2, 18.7, 18.8, 18.10, 18.12 (if the Jacobian criterion is known), 18.14, 18.15

Chapter 19: Homological Theory of Regular Local Rings. All but Corollary 19.11 (or go back to pick up the necessary material from Chapter 15).
Exercises: 19.1, 19.2, 19.3, 19.14, 19.15, 19.16

Chapter 20: Free Resolutions and Fitting Invariants. Through 20.15 (Hilbert-Burch theorem).
Exercises: 20.13, 20.15, 20.17, 20.22, 20.23

Chapter 21: Duality, Canonical Modules, and Gorenstein Rings. As time permits.
Exercises: 21.1, 21.6, 21.7, 21.8, 21.11, 21.18

Acknowledgements

I am personally grateful to Irving Kaplansky for initiating me into the subject of commutative algebra, and to David Buchsbaum for teaching me how to use it. I learned about the geometric content of the ideas first in the lectures of David Mumford, and later in collaborations and less formal contacts with many people, principally Michael Artin, Joe Harris, Bernard Teissier, and Antonius Van de Ven. I learned about Gröbner bases and their applications from Dave Bayer, Frank-Olaf Schreyer, and Michael Stillman. Mel Hochster graciously shared with me some of his notes on commutative algebra; these influenced in particular some of my treatment of the Cohen structure theory. Although, in true textbook style, their ideas (and those of many others) are reproduced without much attribution, I am deeply grateful to all these teachers and collaborators.

This book grew from courses I have taught over the years, most recently at Harvard and Brandeis. I have made use, with only silent thanks, of comments and remarks that my listeners have contributed along the way. A few of my students have been of tremendous help to me in providing corrections and in some cases simplifications and improvements to my arguments; in particular I want to thank Nick Chavdarov, Irena Peeva, and Vesselin Gasharov for extraordinary help of this sort. Keith Pardue wrote his thesis with me on generic initial ideals and related topics, and the treatment given in Chapter 15 owes much to his insights. He and Irena Peeva also helped me greatly with proofreading.

A number of people have used parts of these notes in their own lectures, and have provided me with feedback. I particularly want to thank Joe Harris, Ray Heitmann, Tony Iarrobino, Sheldon Katz, and Steve Kleiman for this sort of help. I also want to thank Frans Oort who read a large part of the book and made many valuable suggestions.

0

Elementary Definitions

For the sake of establishing a common language, this chapter introduces some notation and elementary definitions such as would appear in many undergraduate algebra courses.

Following the usage introduced by Paul Halmos we shall write "iff" for "if and only if." We use the symbol $\subset$ to mean "contained in or equal to," and write $\subsetneqq$ when equality is not an option. We write $\cong$ for isomorphism, but often use $=$ when the isomorphism is canonical.

0.1 Rings and Ideals

A **ring** is an abelian group R with a multiplication operation $(a, b) \mapsto ab$ and an "identity element" 1, satisfying, for all $a, b, c \in R$:

$$\begin{aligned} a(bc) &= (ab)c && \text{(associativity)} \\ a(b+c) &= ab+ac && \\ (b+c)a &= ba+ca && \text{(distributivity)} \\ 1a &= a1 = a && \text{(identity).} \end{aligned}$$

A ring R is **commutative** if, in addition, $ab = ba$ for all $a, b \in R$. Nearly every ring treated in this book is commutative, and we shall generally omit the adjective.

A **unit** (or **invertible element**) in a ring R is an element u such that there is an element $v \in R$ with $uv = 1$. Such a v is unique. It is denoted u^{-1}, and called the **inverse** of u. A **field** is a ring in which every nonzero

element is invertible. We write $\mathbf{Z}, \mathbf{Q}, \mathbf{R}$, and $\mathbf{C}$ respectively for the ring of integers and the fields of rational, real, and complex numbers.

A **zerodivisor** in R is a nonzero element $r \in R$ such that there is a nonzero element $s \in R$ with $rs = 0$. A nonzero element that is not a zerodivisor is a **nonzerodivisor**.

An **ideal** in a commutative ring R is an additive subgroup I such that if $r \in R$ and $s \in I$, then $rs \in I$. An ideal I is said to be generated by a subset $\mathcal{S} \subset R$ if every element $t \in I$ can be written in the form

$$t = \sum_1^n r_i s_i \quad \text{with } r_i \text{ in } R \text{ and } s_i \text{ in } \mathcal{S}.$$

We shall write $(\mathcal{S})$ for the ideal generated by a subset $\mathcal{S} \subset R$; if $\mathcal{S}$ consists of finitely many elements $s_1, \ldots, s_n$, then we usually write $(s_1, \ldots, s_n)$ in place of $(\mathcal{S})$. By convention, the ideal generated by the empty set is 0. An ideal is **principal** if it can be generated by one element.

An ideal I of a commutative ring R is **prime** if $I \neq R$ (we usually say that I is a **proper ideal** in this case) and if $f, g \in R$ and $fg \in I$ implies $f \in I$ or $g \in I$. Equivalently, I is prime if for any ideals J, K with $JK \subset I$ we have $J \subset I$ or $K \subset I$. It follows by induction on n that if I is prime and contains a product of ideals (or even a product of sets) $J_1 J_2 \cdots J_n$, then I contains one of the J_i. The ring R is called a **domain** if 0 is prime. A **maximal ideal** of R is a proper ideal P not contained in any other proper ideal. If $P \subset R$ is a maximal ideal, then R/P is a field, so P is prime. For reasons explained in Chapter 2, R is called a **local ring** if P is the unique maximal ideal. We sometimes indicate this by saying that (R, P) is a local ring.

An element $h \in R$ is **prime** if it generates a prime ideal—equivalently, h is prime if h is not a unit, and whenever h divides a product fg, then h divides f or h divides g.

A **ring homomorphism**, or **ring map**, from a ring R to a ring S is a homomorphism of abelian groups that preserves multiplication and takes the identity element of R to the identity element of S. Generally we shall omit the adjective "ring" when it is clear from context. A subring of S is a subset closed under addition, subtraction, and multiplication, and containing the identity element of S.

If R and S are rings, then the **direct product** $R \times S$ is the set of ordered pairs (a, b) with $a \in R$ and $b \in S$ made into a ring by defining the operations componentwise:

$$\begin{aligned} (a, b) + (a', b') &= (a + a', b + b') \\ (a, b)(a', b') &= (aa', bb'). \end{aligned}$$

Note that the map $a \mapsto (a, 0)$ makes R a subset of $R \times S$, and similarly with S; as subsets of $R \times S$ we have $RS = 0$. Consider the elements $e_1 = (1, 0)$

and $e_2 = (0, 1)$ of $R \times S$. They are **idempotent** in the sense that $e_1^2 = e_1$ and $e_2^2 = e_2$. Furthermore, they are **orthogonal idempotents** in the sense that $e_1 e_2 = 0$. They are even a **complete set of orthogonal idempotents** in the sense that, in addition, $e_1 + e_2 = 1$. Quite generally, if $e_1, \ldots, e_n$ is a complete set of orthogonal idempotents in a commutative ring R, then $R = Re_1 \times \cdots \times Re_n$ is a direct product decomposition.

If R is a commutative ring, then a **commutative algebra** over R (or **commutative *R*-algebra**) is a commutative ring S together with a homomorphism $\alpha : R \to S$ of rings. We usually suppress the homomorphism α from the notation, and write rs in place of $\alpha(r)s$ when $r \in R$ and $s \in S$. Any ring is a $\mathbf{Z}$-algebra in a unique way. A more interesting example of an R-algebra is a polynomial ring $S = R[x_1, \ldots, x_n]$ in finitely many variables. A **subalgebra** of S is a subring S' that contains the image of R. A **homomorphism of *R*-algebras** $\varphi : S \to T$ is a homomorphism of rings such that $\varphi(rs) = r\varphi(s)$ for $r \in R, s \in S$. Given an ideal $I \subset S$ we shall often be interested in its preimage in R. We shall sometimes denote this preimage by $R \cap I$, even though R need not be a subset of S.

The commutative algebras that are of greatest interest to us—the ones of which the reader should think when we say "let R be a commutative algebra" (or "let R be a ring")—are those of the form $R = S/I$, where S is a polynomial ring over a field or, at a more sophisticated level, over the integers, or the localization of such a ring at a prime ideal (see Chapter 2 for localization).

We establish some terminology about polynomials: If k is a commutative ring, then a **polynomial ring** over k in r variables $x_1, \ldots, x_r$ is denoted $k[x_1, \ldots, x_r]$. (We shall much less frequently be interested in polynomial rings in infinitely many variables.) The elements of k are generally referred to as **scalars**. A **monomial** is a product of variables; its degree is the number of these factors (counting repeats) so that, for example, $x_1^2 x_2^3 = x_1 x_1 x_2 x_2 x_2$ has degree 5. By convention the element 1 is regarded as the empty product —it is the unique monomial of degree 0. A **term** is a scalar times a monomial. Every polynomial can be written uniquely as a finite sum of nonzero terms. If the monomials in the terms of a polynomial f all have the same degree (or if $f = 0$), then f is said to be **homogeneous**. We also use the word **form** to mean homogeneous polynomial.

If k is a field, and $I \subset k[x]$ is an ideal, and $f \in I$ is an element of lowest degree, then Euclid's algorithm for dividing polynomials shows that f divides every element of I. Thus $k[x]$ is a **principal ideal domain**, a domain in which every ideal can be generated by one element.

0.2 Unique Factorization

Let R be a ring. An element $r \in R$ is **irreducible** if it is not a unit and if whenever $r = st$ with $s, t \in R$, then one of s and t is a unit. A ring R

is **factorial** (or a **unique factorization domain**, sometimes abbreviated **UFD**) if R is an integral domain and elements of R can be factored uniquely into irreducible elements, the uniqueness being up to factors which are units (this is the same sense in which factorization in **Z** is unique). Factoriality played an enormous role in the history of commutative algebra, and it will come up many times in this book. Here is an elementary analysis of the condition:

If R is factorial, and if $a_1, a_2, \ldots$ is a sequence of elements such that a_i is divisible by a_{i+1}, then the prime factors of a_{i+1} (counted with multiplicity) are among the prime factors of a_i, so for large i the prime factorization is the same, and a_i, a_{i+1} differ only by a unit. In the language of ideals, any increasing sequence of principal ideals $(a_1) \subset \cdots \subset (a_i) \subset \cdots$ must terminate in the sense that for all large i we have $(a_i) = (a_{i+1})$. This condition is called the **ascending chain condition on principal ideals**.

Furthermore, if R is factorial then the irreducible elements of R are prime, that is, they generate prime ideals. (*Proof*: Suppose R is factorial and r is irreducible. If $st \in (r)$, then $st = ru$ for some element u, and by the uniqueness of factorizations, r must divide one of s and t.)

Conversely, if R has ascending chain condition on principal ideals, then any element of R can be factored into a product of irreducible elements: For suppose $a_1 \in R$ admits no factorization into irreducibles (and is not a unit). As a_1 is not irreducible, it can be factored as bc with neither b nor c a unit. Clearly not both b and c can have factorizations into irreducible elements, or putting them together would result in a factorization of a_1. Say b admits no factorization into irreducibles. Setting $a_2 = b$, we have $(a_1) \subsetneqq (a_2)$. Repeating the argument inductively, we get a nonterminating sequence of principal ideals $(a_1) \subsetneqq (a_2) \subsetneqq \cdots$, contradicting our assumption.

If, in addition, every irreducible element of R is prime, then factorization into products of irreducible elements is unique, so R is factorial. The key step in the proof is to show that if $st = ru \in R$ with r irreducible, then r divides one of s and t. Since (r) is prime, we must have $s \in (r)$ or $t \in (r)$, which amounts to what we want to prove. The remainder of the proof is exactly as in the case of the integers.

Using these ideas, it is easy to show, for example, that any principal ideal domain R is factorial: First, if $(a_1) \subset \cdots \subset (a_i) \subset \cdots$ is an ascending chain of ideals, then the set $\cup_i(a_i)$ is again an ideal. Since R is a principal ideal domain, it can be generated by one element $b \in \cup_i(a_i)$. Of course, then $b \in (a_i)$ for some i, and it follows that $(a_i) = (a_{i+1}) = \cdots$. This proves the ascending chain condition on principal ideals.

To show that an irreducible element $r \in R$ is prime, note that the ideal (r) is a proper ideal, so (by Zorn's lemma or by the ascending chain condition just established) we may find a maximal ideal P containing r. Since P is principal, we may write $P = (p)$ for some $p \in R$, and we see that $r = sp$ for some $s \in R$. Since r is irreducible, s is a unit, so $(r) = P$. Since maximal ideals are prime, this shows that r is prime.

The polynomial ring in any number of variables over a field or, indeed, over any factorial ring, is again factorial. This is proved in most elementary texts using a result called Gauss' lemma. See, for example, Exercise 3.4.

0.3 Modules

If R is a ring, then an ***R*-module** M is an abelian group with an action of R, that is, a map $R \times M \to M$, written $(r, m) \mapsto rm$, satisfying for all $r, s \in R$ and $m, n \in M$:

$$\begin{aligned} r(sm) &= (rs)m && \text{(associativity)} \\ r(m+n) &= rm + rn && \\ (r+s)m &= rm + sm && \text{(distributivity, or bilinearity)} \\ 1m &= m && \text{(identity).} \end{aligned}$$

The R-modules we shall be most interested in are the ideals I and the corresponding factor rings R/I; but many others intervene in the study of these.

If M is an R-module, we shall write **ann *M*** for the annihilator of M; that is,

$$\operatorname{ann} M = \{r \in R | rM = 0\}.$$

For example, ann $R/I = I$.

It is convenient to generalize this relation. If I and J are ideals of R, we write $(I : J) = \{f \in R | fJ \subset I\}$ for the **ideal quotient**. (The notation is supposed to suggest division, which it represents in case $I = (i)$, $J = (ij)$, and i is a nonzerodivisor.) It is useful to extend this notion to submodules M, N of an R-module P, and write $(M : N) = \{f \in R | fN \subset M\}$. If $I \subset R$ is an ideal and $M \subset P$ is a submodule, then we occasionally write $(M : I)$ or $(M :_P I)$ for the submodule $\{p \in P | Ip \subset M\}$.

A homomorphism (or map) of R-modules is a homomorphism of abelian groups that preserves the action of R. We say that a homomorphism is a **monomorphism** (or an **epimorphism** or an **isomorphism**) if it is an injection (or surjection or bijection) of the underlying sets. The inverse map to an isomorphism is automatically a homomorphism.

If M and N are R-modules, then the **direct sum** of M and N is the module $M \oplus N = \{(m, n) | m \in M, n \in N\}$ with the module structure $r(m, n) = (rm, rn)$. There are natural inclusion and projection maps $M \subset M \oplus N$ and $M \oplus N \to M$ given by $m \mapsto (m, 0)$ and $(m, n) \mapsto m$ (and similarly for N). These maps are enough to identify a direct sum: That is, M is a **direct summand** of a module P iff there are homomorphisms $\alpha : M \to P$ and $\sigma : P \to M$ whose composition $\sigma\alpha$ is the identity map of M; then $P \cong M \oplus$ (ker σ). The simplest modules are the direct sums of copies of R: These are called free R-modules.

Similar considerations hold for the direct sum of any finite set of modules, but for infinite sets of modules $\{M_i\}_{i \in I}$ we must distinguish the **direct**

product $\Pi_i M_i$, whose elements are tuples $(m_i)_{i\in I}$, from the **direct sum**, $\oplus_i M_i \subset \Pi_i M_i$, consisting of those tuples (m_i) such that all but finitely many m_i are 0.

A **free *R*-module** is a module that is isomorphic to a direct sum of copies of R. We usually write R^n for the direct sum of n copies of R, and think of it as a free module with a given basis, namely the set of "coordinate vectors" $(1,0,\ldots,0)$, $(0,1,0,\ldots,0)$, ..., $(0,\ldots,0,1)$. If M is a finitely generated free module, that is $M \cong R^n$ for some n, then the number n is an invariant of M (in the case when R is a field this is just the dimension of M as a vector space). It is called the **rank** of M. For a somewhat unusual proof that the rank is well defined, see Corollary 4.5.

If A, B, and C are R-modules, and $\alpha : A \to B$, $\beta : B \to C$ are homomorphisms, then a pair of homomorphisms

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} C$$

is **exact** if the image of α is equal to $\ker \beta$, the kernel of β. In general, a sequence of maps between modules like

$$0 \to A \to B \to C \to D$$

is exact if each pair of consecutive maps is exact.

For example, a **short exact sequence** is a sequence of maps

$$0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} C \to 0$$

such that each pair of consecutive maps is exact; that is, such that α is an injection, β is a surjection, and the image of α is the kernel of β. The short exact sequence is **split** iff there is a homomorphism $\tau : C \to B$ such that $\beta\tau$ is the identity map of C; then $B \cong A \oplus C$. (*Reason*: If a map τ with the desired property exists, then $\operatorname{im}\tau$, the image of τ, is disjoint from the image of α, and together they generate B, so $B = \alpha(A) \oplus \tau(C)$. But $\alpha(A) \cong A$ and $\tau(C) \cong C$.) Equivalently, the sequence is split iff there exists a homomorphism $\sigma : B \to A$ such that $\sigma\alpha$ is the identity map of A. (*Reason* for the equivalence: Given τ such that $\beta\tau = 1$, set $\sigma' = 1 - \tau\beta : B \to B$. Since $\beta\sigma' = \beta - \beta\tau\beta = \beta - 1\beta = 0$, the image of σ' is contained in the image of α, so we may factor σ' as $\sigma' = \alpha\sigma$ for some map $\sigma : B \to A$. For any $a \in A$ we have $\alpha(\sigma\alpha(a)) = \alpha\sigma(\alpha(a)) = \sigma'\alpha(a) = \alpha(a) - \tau\beta\alpha(a) = \alpha(a)$, and since α is an injection, this implies $\sigma\alpha(a) = a$ so $\sigma\alpha$ is the identity of A. Conversely, given a map σ with $\sigma\alpha = 1$, a dual path leads back to a suitable map τ.)

Here are three common examples that may help make these things clear:

1. If M_1 and M_2 are submodules of a module M, and $M_1 + M_2 \subset M$ is the submodule they generated, then the two inclusion maps combine to give a map $M_1 \cap M_2 \to M_1 \oplus M_2$, and with the "difference" map

$M_1 \oplus M_2 \to M_1 + M_2$ given by $(m_1, m_2) \mapsto m_1 - m_2$, this gives a short exact sequence

$$0 \to M_1 \cap M_2 \to M_1 \oplus M_2 \to M_1 + M_2 \to 0,$$

as the reader may easily check. The case of vector spaces is probably already familiar, and this case is no different.

2. If R is a ring, $I \subset R$ an ideal, and $a \in R$ an element, then R/I maps onto $R/(I + (a))$. The kernel is generated by the class of a modulo I. Since the kernel is generated by just one element, it has the form R/J for some ideal J; in fact, J is the annihilator of a modulo I, that is, $J = (I : a)$. Putting this together, we see that there is an exact sequence

$$0 \to R/(I : a) \xrightarrow{a} R/I \to R/(I + (a)) \to 0,$$

where the element a over the left-hand map indicates that it is multiplication by a.

3. One way to specify an R-module is by giving "generators and relations": For example, if we say that a module has one generator g and relations $f_1 g = f_2 g = \cdots = f_n g = 0$, for some elements $f_1, \ldots, f_n \in R$, then the module is $R/(f_1, \ldots, f_n)$. Here is an exact sequence view:

An element m of a module M corresponds to a homomorphism from R to M, sending 1 to m. Thus, giving a set of elements $\{m_\alpha\}_{\alpha \in A} \in M$ corresponds to giving a homomorphism φ from a direct sum $G := R^A$ of copies of R, indexed by A, to M, sending the αth basis element to m_α. If the m_α generate M, then φ is a surjection.

The relations on the m_α are the same as elements of the kernel of the map $G \to M$. A set of relations $\{n_\beta\}_{\beta \in B} \in G$ corresponds to a homomorphism ψ from a free module $F := R^B$ to the kernel of φ. The m_α generate M and the n_β generate the kernel —that is, M may be described as the module with generators $\{m_\alpha\}_{\alpha \in A}$ and relations $\{n_\beta\}_{\beta \in B}$ —iff the sequence

$$F \to G \to M \to 0$$

is exact. This sequence is usually called a **free presentation** of M. In case A and B are finite sets, so that each of F and G is a finitely generated free module over R, it is called a **finite free presentation**. A module M is **finitely generated** if there exists a finite set of elements that generate M, and **finitely presented** if it has a finite free presentation.

Part I

Basic Constructions

1
Roots of Commutative Algebra

This chapter describes the origins of commutative algebra and follows its development through the landmark papers published by David Hilbert in 1890 and 1893. Three major strands of nineteenth-century activity lie behind commutative algebra and are still its primary fields of application: number theory, algebraic geometry (the algebraic aspect really begins with Riemann's "function theory"), and invariant theory. We shall say a little about developments in each.

Advice for the beginner: A complete understanding of this chapter would require more background than is necessary for the rest of this book, and you should feel free to read lightly over the more difficult parts. Most of the topics treated here are taken up again later, with greater generality and in greater detail. In order to go on, you need to master only Theorem 1.2 and its Corollaries 1.3, 1.4, the definition of a graded ring in Section 1.5, and Theorem 1.11, the fact that the Hilbert function becomes a polynomial (this last is not actually needed until Chapter 12).

1.1 Number Theory

Interest in the objects that we now associate with commutative algebra probably first arose in number theory. After $\mathbf{Z}, \mathbf{Q}, \mathbf{R}$, and $\mathbf{C}$, perhaps the very first ring of interest was the ring of "Gaussian integers" $\mathbf{Z}[i]$, with $i^2 = -1$, introduced and exploited by Gauss in his 1828 paper on biquadratic residues. Gauss proved that the elements of $\mathbf{Z}[i]$ admit unique

factorization into prime elements, just as is the case for ordinary integers, and he exploited this unique factorization to prove results about the ordinary numbers.

Number theorists soon appreciated how useful it was to adjoin solutions of polynomial equations to **Z**, and they found that in many ways the enlarged rings behaved much like **Z** itself. Euler, Gauss, Dirichlet, and Kummer all used this idea for the rings $\mathbf{Z}[\zeta]$, with ζ a root of unity, to prove some special cases of Fermat's last theorem (the insolubility in integers of the equation $x^n + y^n = z^n$). Around 1847, Lamé thought he had a proof in general based on this method, but Liouville was quick to point out problems. Kummer, who already knew the error, did succeed in proving the result for $n < 100$ in 1851. The idea behind these proofs is rather obvious, and obviously attractive: If ζ is an nth root of -1, then $x^n + y^n = \Pi_i(x - \zeta^{2i+1}y)$. If $\mathbf{Z}[\zeta]$ has unique factorization into primes, it is profitable to compare the factorization of $x^n + y^n$ as $\Pi_i(x - \zeta^{2i+1}y)$ with the factorization as z^n. It is a plausible conjecture that Fermat's unreported "proof" (the one that was too long to fit in the margin of his copy of Diophantus' book) was also based on this idea.

The problem with these proofs is that for most n the ring $\mathbf{Z}[\zeta]$ does not have unique factorization (the first example is $n = 23$). The search for some generalization of unique factorization that might be used instead guided a large proportion of early commutative algebra. Most significant for modern algebra is surely Dedekind's introduction of **ideals** of a ring; the name comes from the view that they represent "ideal" (that is to say, "not real") elements of the ring. The search for unique factorization culminated in two major theories, which we shall describe later: Dedekind's unique factorization of ideals into prime ideals in the rings we now call Dedekind domains; and Kronecker's theory of polynomial rings and Lasker's theory of primary decomposition in them.

Dedekind's idea was to represent an element $r \in R$ by the ideal (r) of its multiples; arbitrary ideals might thus be regarded as ideal elements. The ideal (r) determines the element r only up to multiples by **units** u of R. Since "unique prime factorization" is only unique up to unit multiples anyway, this is just right for generalizing prime factorization. Dedekind sought and found conditions under which a ring has unique factorization of ideals into prime ideals—he showed that this occurs for the ring of all integers in any number field. Dedekind made these definitions, together with the definition of a ring itself, in a famous supplement to later editions (after 1871) of Dirichlet's book on number theory.

Dedekind's ideas restored a kind of unique prime factorization—of ideals in terms of prime ideals—to the rings with which Kummer was dealing; unfortunately, they did not rescue the proof of Fermat's last theorem. (Perhaps this was fortunate after all, given the immense amount of mathematics that this area of number theory has spawned.) The rings for which

Dedekind's theory works are now called *Dedekind domains* in his honor; they are treated in Chapter 11 of this book.

Around the same time, Kronecker (who was incidentally Kummer's student; Dedekind had been Gauss' student) took a step that led to a different generalization of unique factorization. In his memoir [1881], he put the notion of "adjoining a root of a polynomial equation $f(x) = 0$ to a field k" on a firm footing by introducing the idea of the polynomial ring $k[x]$ in an "indeterminate" x over k; the desired ring is then $k[x]/(f(x))$, and the image of x in this ring is the desired root. He introduced a theory for these polynomial rings equivalent to Dedekind's theory of ideals. What we would call an ideal in the polynomial ring, he called a "modular system" or "module." (The origin of the term is an older usage, which survives today in statements such as, "7 is congruent to 3 modulo 4.") There is no way to factorize ideals in polynomial rings multiplicatively, as in Dedekind's theory, but Lasker [1905] showed how to generalize unique factorization into **primary decomposition** (treated in Chapter 3 of this book).

Both Dedekind's and Lasker's theories were thoroughly reformulated and axiomatized by Emmy Noether in the 1920s, initiating the modern development of commutative algebra.

1.2 Algebraic Curves and Function Theory

> L'algèbre n'est qu'une géométrie écrite; la géométrie n'est qu'une algèbre figurée.
>
> (Algebra is but written geometry; geometry is but drawn algebra.)
>
> —Sophie Germain (1776–1831)

The study of algebraic curves in the early nineteenth century is in retrospect very closely related to commutative algebra, but the connection hardly began to appear until the 1870s and 1880s. Conics had of course been studied since antiquity. The work of Fermat and Descartes on coordinate geometry made it possible to speak of the (real) plane curves of any degree represented by algebraic equations, and these were studied intensely in the eighteenth century (for example, Isaac Newton classified real plane cubics (curves in $\mathbf{R}^2$ defined by the vanishing of a polynomial $f(x, y)$ of degree 3) into families—there are more than 90; and MacLaurin showed in 1720 that a plane curve of degree d could have at most $(d-1)(d-2)/2$ nodes), along with some curves and surfaces in three-space. However, the ideas necessary for associating rings to these objects were entirely absent. Indeed, until the introduction of complex numbers by Gauss and others, early in the nineteenth century, a close connection of the kind explained later in this chapter was out of reach.

About 1860 the work of Abel, Jacobi, and Riemann made an entirely new view of algebraic curves possible. Clebsch, around 1864, was the first to apply Riemann's ideas directly to plane curves. The new emphasis was mostly on the field of meromorphic functions on a curve. Kronecker, Weierstrass, Dedekind, and Weber discovered in the period from 1875 to 1882 that many of the recently developed algebraic techniques for handling number fields could be applied to these geometrically defined fields; they pioneered what was then called the "arithmetic approach to function theory." This approach continued to develop through the end of the nineteenth century and is well represented by Hensel and Landsberg [1902]. The work of Dedekind and Weber might have been described at the time as the application of ideas from number theory to problems from analysis. It seems now to be the real beginning of the interaction of geometry with commutative algebra, the central theme of this book.

1.3 Invariant Theory

> As all roads lead to Rome so I find in my own case at least that all algebraic inquiries, sooner or later, end at the Capitol of modern algebra over whose shining portal is inscribed the Theory Of Invariants.
>
> —J.J. Sylvester [1864, p. 380]

The work on ideals done in the 1880s, both in number-theoretic and function-theoretic contexts, seems a trifle quaint to modern readers; but the work of Hilbert just a few years later seems quite modern. In two extraordinary papers [1890, 1893], which are still a pleasure to read, Hilbert greatly advanced the theory of ideals in polynomial rings. Hilbert's motivation comes from a subject we have not yet mentioned: the theory of invariants. We shall sketch a little of this theory. For systematic modern accounts, see Fogarty [1969], Kraft [1985], and Sturmfels [1992].

Especially after the introduction of projective coordinates by Plücker around 1830, people became interested in the geometric properties of plane curves that were invariant under certain classes of transformations. One way to express such an invariant property is to give some sort of function that associates to a geometric configuration a number that is independent of the choice of coordinates.

As time went on, mathematicians realized that the invariance under choice of coordinates was really the invariance under an action of a group, typically the special linear group $\mathrm{SL}_n(k)$ of $n \times n$ matrices of determinant 1 with entries in k, or the general linear group $\mathrm{GL}_n(k)$ of all invertible matrices with entries in R, or a finite group. The functions studied were mostly polynomial functions of quantities defining the geometric objects, such as the coefficients of the equations of algebraic plane curves. Thus the general

problem of invariant theory came to be the following: Given a "nice" action of a group G as automorphisms of a polynomial ring $S = k[x_1, \dots, x_r]$, find the elements of S that are left invariant by G. The set of invariant elements, written S^G, forms a subalgebra of S. In many interesting cases people saw that they could find a finite set of invariants generating the ring S^G, and in this way they could describe all the invariants in finite terms.

Invariant theory has always been a subject of examples, and the following is a central one.

Example 1.1. Let $S = k[x_1, \dots, x_r]$ be the polynomial ring, and let Σ be the symmetric group of all permutations of $\{1, \dots, r\}$. The group Σ acts on S as follows: If $\sigma \in \Sigma$ and $f \in S$, we define

$$(*) \qquad \sigma(f)(x_1, \dots, x_r) = f(x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(r)}).$$

The group Σ then acts as a group of k-algebra automorphisms of S. The set of *invariants*

$$S^\Sigma := \{f \in S | \sigma(f) = f\},$$

which in this case is called the **ring of symmetric functions**, is therefore a subring of S. It obviously contains the **elementary symmetric functions**

$$\begin{aligned} f_1(x_1, \dots, x_r) &:= x_1 + \cdots + x_r, \\ f_2(x_1, \dots, x_r) &:= \sum_{1 \le i < j \le r} x_i x_j, \\ \dots\dots \\ f_r(x_1, \dots, x_r) &:= x_1 \cdot x_2 \cdot \cdots \cdot x_r. \end{aligned}$$

In fact, S^Σ is generated as a k-algebra by $f_1, \dots, f_r$, and every symmetric function can be written uniquely as a polynomial in the f_i (see Exercise 1.6 for a proof). Thus S^Σ is isomorphic to a polynomial ring $k[y_1, \dots, y_r]$ by the map sending y_i to f_i.

A great deal of late nineteenth-century work was devoted to the problem of finding finite systems of generators for rings of invariants in similarly explicit cases. For example, if we let $F = x_0 s^d + x_1 s^{d-1} t + \dots + x_d t^d$ be the "general" form of degree d in variables s, t, then for $a, b, c, d \in \mathbf{C}$ a substitution $s = as' + bt'$, $t = cs' + dt'$ leads to an expression of F in terms of monomials in s' and t' with new coefficients $x'_0, \dots, x'_d$ that are linear combinations of $x_0, \dots, x_d$. Restricting to invertible substitutions of this type with determinant 1, we get an action of the group $SL_2(\mathbf{C})$ on the polynomial ring $\mathbf{C}[x_0, \dots, x_d]$. The "Problem of invariants of binary forms of degree d" is to find the invariants of this action. This remains a hard problem: Systems of generators are still not known, when d is large. The **fundamental problem of invariant theory** was the problem of the existence of finite systems of generators.

Hilbert solved this problem in a spectacular series of papers from 1888 to 1893, showing that the ring of invariants is finitely generated in a wide range of cases, including the ones above.[1] The proof that we shall soon give parallels Hilbert's, though we have modernized it slightly. Hilbert's proof is quite nonconstructive and is said to have provoked Paul Gordan, the reigning "king of invariants," to remark: "This is not mathematics but theology!" Hilbert returned to the problem in a later paper [1893] and gave a proof that is constructive (see Sturmfels [1993] for a modern discussion). Gordan, for his part, was quick to understand and appreciate Hilbert's new idea; he simplified Hilbert's nonconstructive proof in a paper of his own, and remarked, "I have convinced myself that Theology also has its advantages." (*Nachrichten König. Ges. der Wiss. zu Gött.*, 1899, 240–242; the story is from Kline [1972], p. 930. We shall give what is essentially Gordan's proof in Exercise 15.15.) Hilbert's work is often said to have killed invariant theory by solving its central problem. But mathematics seems to be immortal. After a period of relatively little activity, invariant theory has enjoyed a resurgence in our day, as the books quoted above indicate; and it has a whole new branch, geometric invariant theory, of which we shall say a little after we introduce the Nullstellensatz.

Aside from the invariant theory, Hilbert proved four major results in the papers of 1890 and 1893: the basis theorem (which leads directly to the finite generation of invariants), the "theorem of zeros" (traditionally called by its German name, the Nullstellensatz), the polynomial nature of what we call the Hilbert function, and the syzygy theorem. These results have played an enormous role in determining the shape of commutative algebra. There seems no better introduction to the subject than to discuss them in turn.

1.4 The Basis Theorem

The first step in Hilbert's proof of the finiteness of invariants was the Basis Theorem: If R is a polynomial ring in finitely many variables over a field or over the ring of integers, then every ideal in R can be generated by finitely

[1]Hilbert remained interested in the problem afterward. In 1900 he gave an address to the International Congress of Mathematicians containing a list of problems that has since become quite celebrated. The fourteenth problem asks whether there is a finite basis for the invariants of any linear group acting on a polynomial ring by linear change of coordinates, or for still more general subrings. The first counterexample was found by Nagata in 1959. But a closely related problem first studied by Zariski remains central. Perhaps its most interesting avatar is the problem of the "finite generation of the canonical ring of a variety of general type," whose solution in dimension three was one of the key steps in the work for which Mori won a Fields medal in 1986.

many elements (the word "basis" at the time simply meant "generators"). This key property is now named not after Hilbert, but after Emmy Noether, who realized its full importance. (Interestingly, Noether was a student of Gordan.) Noether showed in [1921] how to use the property as a basic axiom in commutative algebra. In particular, she showed that results such as Lasker's "primary decomposition," which had seemed to rest on the innermost nature of polynomial rings, could be derived very simply with just this axiom. See Exercise 1.2 for a central example.

We say, then, that a ring R is **Noetherian** if every ideal of R is finitely generated; it is easy to see that this is equivalent to the **ascending chain condition on ideals of *R***, which says that every strictly ascending chain of ideals must terminate. (*Proof*: If $I \subset R$ is an ideal, then by successively choosing elements f_i of I, we get a chain of ideals $(f_1) \subset (f_1, f_2) \subset \cdots$ that can be made to ascend forever unless one of them is equal to I. Thus if R has ascending chain condition, then I is finitely generated. Conversely, if $I_1 \subsetneqq I_2 \subsetneqq \cdots$ is a strictly ascending chain of ideals of R, and the ideal $\cup_i I_i$ has a finite set of generators, then these generators must all be contained in one of the I_j; and thus $I_j = I$, and the ascending chain terminates at I_j.) The ascending chain condition may be restated by saying that every collection of ideals in R has a maximal element. See Exercise 1.1 for Hilbert's original statement.

For example, any field is Noetherian (the only ideals are 0 and the whole field) and the ring $\mathbf{Z}$ of integers is Noetherian (each ideal is generated by a single integer, the greatest common divisor of the elements of the ideal). Hilbert originally showed that a polynomial ring in n variables over a field or over the ring of integers is Noetherian. The modern version is somewhat more general. (Hilbert's version is contained in Corollary 1.3.)

Theorem 1.2 (Hilbert Basis Theorem). *If a ring R is Noetherian, then the polynomial ring $R[x]$ is Noetherian.*

The following notion will be useful in the proof and later: If $f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in R[x]$, with $a_n \neq 0$, we define the **initial term** of f to be $a_n x^n$, and we define the **initial coefficient** of f to be a_n.

Proof. Let $I \subset R[x]$ be an ideal; we shall show that I is finitely generated. Choose a sequence of elements $f_1, f_2, \ldots \in I$ as follows: Let f_1 be a nonzero element of least degree in I. For $i \geq 1$, if $(f_1, \ldots, f_i) \neq I$, then choose f_{i+1} to be an element of least degree among those in I but not in $(f_1, \ldots, f_i)$. If $(f_1, \ldots, f_i) = I$, stop choosing elements.

Let a_j be the initial coefficient of f_j. Since R is Noetherian, the ideal $J = (a_1, a_2, \ldots)$ of all the a_i produced is finitely generated. We may choose a set of generators from among the a_i themselves. Let m be the first integer such that $a_1, \ldots, a_m$ generate J. We claim that $I = (f_1, \ldots, f_m)$.

In the contrary case, our process chose an element f_{m+1}. We may write $a_{m+1} = \sum_{j=1}^{m} u_j a_j$, for some $u_j \in R$. Since the degree of f_{m+1} is at least as great as the degree of any of the $f_1, \ldots, f_m$, we may define a polynomial $g \in R$ having the same degree and initial term as f_{m+1} by the formula

$$g = \sum_{j=1}^{m} u_j f_j x^{\deg f_{m+1} - \deg f_j} \in (f_1, \ldots, f_m).$$

The difference $f_{m+1} - g$ is in I but not in $(f_1, \ldots, f_m)$, and has degree strictly less than the degree of f_{m+1}. This contradicts the choice of f_{m+1} as having minimal degree. The contradiction establishes our claim. □

The basis theorem can be applied to any finitely generated algebra.

Corollary 1.3. *Any homomorphic image of a Noetherian ring is Noetherian. Furthermore, if R_0 is a Noetherian ring, and R is a finitely generated algebra over R_0, then R is Noetherian.*

Proof. Given an ideal I in R/J, with R Noetherian, the preimage of I in R is finitely generated, and the images of its generators generate I.

Since R is a finitely generated algebra over R_0, R is a homomorphic image of $S := R_0[x_1, \ldots, x_r]$ for some r. Using Theorem 1.2 and induction on r, we see that S is Noetherian. Since a homomorphic image of a Noetherian ring is Noetherian, we are done. □

We shall need a more general definition in the sequel, and we make it now: An R-module M is **Noetherian** if every submodule of N is finitely generated. By the same argument as above, this is equivalent to the condition that M has ascending chain condition on submodules, or again that every collection of submodules of M has a maximal element. The importance of Noetherian modules comes from the following observation:

Proposition 1.4. *If R is a Noetherian ring and M is a finitely generated R-module, then M is Noetherian.*

Proof. Suppose that M is generated by $f_1, \ldots, f_t$, and let N be a submodule. We shall show that N is finitely generated by induction on t.

If $t = 1$, then the map $R \to M$ sending 1 to f_1 is surjective. The preimage of N is an ideal, which is finitely generated since R is Noetherian. The images of its generators generate N.

Now suppose $t > 1$. The image $\bar{N}$ of N in M/Rf_1 is finitely generated by induction. Let $g_1, \ldots, g_s$ be elements of N whose images generate $\bar{N}$. Since $Rf_1 \subset M$ is generated by one element, its submodule $N \cap Rf_1$ is finitely generated, say by $h_1, \ldots, h_r$.

We shall show that the elements $h_1, \ldots, h_r$ and $g_1, \ldots, g_s$ together generate N: Given $n \in N$, the image of n in $\bar{N}$ is a linear combination of the

images of the g_i; so subtracting the corresponding linear combination of the g_i from n itself, we get an element of $N \cap Rf_1$, that is a linear combination of the h_i by hypothesis. This shows that n is a linear combination of the g_i and h_i. □

1.4.1 Finite Generation of Invariants

Hilbert's original application, the existence of finite bases of invariants, is a good illustration of the power of the basis theorem. We shall abstract what we need about the rings of invariants Hilbert considered, but for the interested reader, here are some details.

Let k be a field of characteristic 0 (Hilbert would have taken $\mathbf{C}$) and let G be a finite group or one of the "linear groups" $\mathrm{SL}_n(k)$ or $\mathrm{GL}_n(k)$. The ideas we shall present can be generalized to a much wider class of groups and fields, and the type of actions treated can be greatly extended, but the cases we shall treat remain central examples. See Kraft [1985].

Suppose that $S = k[x_1, \ldots, x_r]$ is a polynomial ring, and that G is represented as a group of linear transformations of the vector space of linear forms of S —that is, we are given a homomorphism of groups $G \to \mathrm{GL}_r(k)$, where we regard the latter group as the group of invertible linear transformations of the vector space with basis $x_1, \ldots, x_r$. If G is $\mathrm{SL}_n(k)$ or $\mathrm{GL}_n(k)$, then we restrict attention to the cases where the representation is **rational** in the following sense: Regarding elements of G as matrices, we require that the matrix by which an element $g \in G$ acts has entries that are rational functions in the entries of g. We extend the action of an element $g \in G$ to all of S by setting $g(f)(x_1, \ldots, x_n) = f(g^{-1}(x_1), \ldots, g^{-1}(x_r))$, and G becomes in this way a group of automorphisms of S. An invariant of G is a polynomial left invariant by each element of G, and the set S^G of invariants is a subring of S.

Hilbert used two basic facts about the ring of invariants $R = S^G$ in the cases he considered. First, R may be written as a direct sum of the vector spaces R_i consisting of homogeneous forms of degree i that are invariant under G. This situation will occur so frequently, and plays such an important part in commutative algebra generally, that we pause here to abstract it.

1.5 Graded Rings

A **graded ring** is a ring R together with a direct sum decomposition

$$R = R_0 \oplus R_1 \oplus R_2 \oplus \cdots \quad \text{as abelian groups,}$$

such that

$$R_i R_j \subset R_{i+j} \quad \text{for } i, j \geq 0.$$

A **homogeneous element** of R is simply an element of one of the groups R_i, and a **homogeneous ideal** of R is an ideal that is generated by homogeneous elements. (Note that since the sum of homogeneous elements of different degrees is not homogeneous, homogeneous ideals contain lots of nonhomogeneous elements.) If $f \in R$, there is a unique expression for f of the form

$$f = f_0 + f_1 + \cdots \quad \text{with } f_i \in R_i \text{ and } f_j = 0 \quad \text{for } j;$$

the f_i are called the **homogeneous components** of f. (One can enlarge these definitions to allow components of negative degrees: We shall sometimes call the result a **Z**-graded ring. More generally, one can imagine a ring graded by any semigroup with identity; we shall occasionally meet $\mathbf{Z}^n$-graded rings in the sequel, and $\mathbf{Z}/(2)$-graded rings are also important.) Although it is the most important ideal of R, the ideal consisting of all elements of degree greater than 0 is called the **irrelevant ideal** (the reason will become clear when we come to the connection with projective geometry), written R_+.

The simplest example of a graded ring is the ring of polynomials $S = k[x_1, \ldots, x_r]$ **graded by degree:** that is, with grading

$$S = S_0 \oplus S_1 \oplus \cdots,$$

where S_d is the vector space of homogeneous polynomials (also called forms) of degree d.

Suppose that I is a homogeneous ideal of a graded ring R, and I is generated by homogeneous elements $f_1, \ldots, f_s$. If $f \in I$ is any homogeneous element, then we can write $f = \sum g_i f_i$ with each g_i homogeneous of degree $\deg g_i = \deg f - \deg f_i$. Indeed, if $f = \sum G_i f_i$ is any expression with $G_i \in R$, then we may take g_i to be the homogeneous component of G_i of degree equal to $\deg f - \deg f_i$; all the other terms in the sum must have cancelled anyway. This apparently innocuous fact about graded rings is actually quite powerful. The ungraded situation is far more complicated; see the remark after Corollary 1.7.

The second fact about invariants that we shall use is that, in the cases we are treating, there is a map of S^G-modules $\varphi : S \to S^G$, which preserves degrees and takes each element of S^G to itself. In case G is a finite group, this is easy: If γ is the number of elements in G, then because k has characteristic 0, γ has an inverse $1/\gamma \in k$, and the "averaging" map φ taking $f \in S$ to $\varphi(f) = (1/\gamma)\sum_{\sigma \in G} \sigma(f)$ has the desired properties. In the case where $G = \mathrm{GL}_n(k)$ or $\mathrm{SL}_n(k)$, acting rationally, φ may be constructed by replacing the sum above with an integral; see Kraft [1985]. Hilbert himself did not know the existence of the map φ in the case of $\mathrm{SL}_n(k)$ and $\mathrm{GL}_n(k)$, and used a map with a weaker property, "Cayley's Ω-process." See Sturmfels [1993, Chapter 4.3].

Hilbert's finiteness result follows at once by taking $R = S^G$ in the following:

Corollary 1.5. *Let k be a field, and let $S = k[x_1, \ldots, x_r]$ be a polynomial ring graded by degree. Let R be a k-subalgebra of S. If R is a summand of S, in the sense that there is a map of R-modules $\varphi : S \to R$ that preserves degrees and takes each element of R to itself, then R is a finitely generated k-algebra.*

Proof. Let $\mathfrak{m} \subset R$ be the ideal generated by the homogeneous elements of R of strictly positive degree. Since S is Noetherian, the ideal $\mathfrak{m}S$ has a finite set of generators, which may be chosen to be homogeneous elements $f_1, \ldots, f_s$ of $\mathfrak{m}$. We shall show that these elements generate R as a k-algebra.

To do this, let R' be the k-subalgebra of S generated by $f_1, \ldots, f_s$, and suppose $f \in R$. We shall show that $f \in R'$ by induction on the degree of f. To start the induction, note that if degree $f = 0$, then $f \in k \subset R'$, as claimed.

Now suppose $\deg f > 0$, so that $f \in \mathfrak{m}$. Since the f_i generate $\mathfrak{m}S$ as an ideal of S, we may write $f = \sum g_i f_i$, where each g_i is a homogeneous form of degree

$$\deg g_i = \deg f - \deg f_i < \deg f.$$

Applying φ, and using the fact that f and the f_i are in R, we get $f = \sum \varphi(g_i) f_i$. Since $\varphi(g_i)$ has lower degree than f, we have $\varphi(g_i) \in R'$ by induction. Thus $f \in R'$ as required. $\square$

For an analysis of the idea behind this surprising proof, see Exercises 1.4 and 1.5.

1.6 Algebra and Geometry: The Nullstellensatz

Gauss' **fundamental theorem of algebra** establishes the basic link between algebra and geometry: It says that a polynomial in one variable over $\mathbf{C}$, an algebra object, is determined up to a scalar factor by the set of its roots (with multiplicities), a geometric object. **Hilbert's Nullstellensatz** extends this link to certain ideals of polynomials in many variables. It is a formal consequence of the fundamental theorem of algebra in the sense that it holds for any algebraically closed field. We shall now sketch this very important connection. (For a more detailed treatment see Fulton [1969] or Cox, Little, and O'Shea [1992].)

A polynomial $f \in k[x_1, \ldots, x_n]$ with coefficients in a field k defines a function, $f : k^n \to k$; the value of f at a point $(a_1, \ldots, a_n) \in k^n$ is obtained by substituting the a_i for the x_i in f. The function defined by f is called a **polynomial function** on the n-dimensional vector space k^n over k, with values in k. If k is infinite, then no polynomial function other than 0 can vanish identically on k^n. (*Reason*: The case of one variable is the statement that a polynomial in one variable can have only finitely many roots, and follows from Euclid's algorithm for division. In the general case we think

of a nonzero polynomial $f(x_1, \ldots, x_n)$ in n variables as a polynomial in $n-1$ variables with coefficients that are polynomials in one variable. By the preceding case we can specialize this one variable to a scalar in such a way that the polynomial remains nonzero, and we are done by induction on the number of variables.)

If follows that if k is infinite, then distinct polynomials define distinct functions. Thus we may regard the polynomial ring $k[x_1, \ldots, x_n]$ as the ring of polynomial functions on k^n. Viewed with its ring of polynomial functions, k^n is usually called **affine n-space** over k, written $\mathbf{A}^n(k)$ or simply $\mathbf{A}^n$. (If k is not algebraically closed, it is useful and customary in algebraic geometry to make a distinction between $\mathbf{A}^n(k)$ and k^n; see Eisenbud and Harris [1992]. This will not concern us here.)

Given a subset $I \subset k[x_1, \ldots, x_n]$, we define a corresponding **algebraic subset** of k^n to be

$$Z(I) = \{(a_1, \ldots, a_n) \in k^n | f(a_1, \ldots, a_n) = 0 \quad \text{for all } f \in I\}.$$

Such algebraic sets are sometimes called an **affine algebraic sets** to distinguish them from the "projective" objects we shall define later.

From the definition of the algebraic subset $Z(I)$, it is clear that I may be replaced by the ideal that it generates in $k[x_1, \ldots, x_n]$ without changing $Z(I)$.

If $X = Z(I)$ is an algebraic set, then an **algebraic subset** $Y \subset X$ is a set of the form $Y = Z(J)$ that happens to be contained in X. An algebraic set is called **irreducible** if it is not the union to two smaller algebraic subsets. Irreducible algebraic sets are called **algebraic varieties** (this name is used by some authors for all algebraic sets, but we shall maintain the distinction).

If $k = \mathbf{R}$ or $k = \mathbf{C}$, then k^r is naturally a topological space (as a product of copies of k), and an algebraic subset $X \subset \mathbf{A}^r$ inherits the subspace topology, called the **classical topology**. But there is another, coarser, topology on X that is defined over any field. Polynomial functions on X will play the role of continuous functions, even when the fields we are working over have no topology, and by analogy with the continuous case it is natural to think of an algebraic subset Y as a **closed** subset of X. Since we obviously have $\cap_i Z(J_i) = Z(\cup_i J_i)$, the intersection of any collection of algebraic subsets is algebraic. Furthermore, if we define $\Pi_{r=1}^n J_i$ to be the set consisting of all products of one function from each J_i, then $\cup_{i=1}^n Z(J_i) = Z(\Pi_{i=1}^n J_i)$, so any finite union of algebraic subsets is algebraic. Thus we may define a topology on X by taking the closed sets to be the algebraic subsets of X. This topology is called the **Zariski topology** in honor of Oscar Zariski, one of the pioneers of work with algebraic varieties over arbitrary fields. The Zariski topology is much coarser than the classical topology when $k = \mathbf{R}$ or $\mathbf{C}$, but it is still quite useful. Some additional information and an important extension of the idea will be found in Exercise 1.24.

There is a sort of inverse to the construction of an algebraic set: Given any set $X \subset k^n$, we define

$$I(X) = \{f \in k[x_1, \ldots, x_n] | f(a_1, \ldots, a_n) = 0 \quad \text{for all } (a_1, \ldots, a_n) \in X\}.$$

It is clear that $I(X)$ is an ideal. A **polynomial function** (or **regular function**) on X is by definition the restriction of a polynomial function on k^n to X. Identifying two polynomial functions if they agree at all the points of X, we get the **coordinate ring** $\boldsymbol{A(X)}$ of X (so called because it is the k-algebra of functions on X generated by the "coordinate functions" x_i). Clearly we have $A(X) = k[x_1, \ldots, x_n]/I(X)$.

Not every homomorphic image $A = k[x_1, \ldots, x_n]/I$ could be the coordinate ring of a set. For suppose an element $f \in A$ satisfies $f^n = 0$. If f were a function on some set X, then because evaluation at a point $p \in X$ is a ring homomorphism, we would have $0 = f^d(p) = f(p)^d$; that is, $f(p)$ is **nilpotent** for all $p \in X$. But the values of f are elements of k, a field; so they are all 0, and f itself is the zero element of $A(X)$. In general, a ring is said to be **reduced** if its only nilpotent element is 0; we have just shown that $A(X)$ is reduced.

It is easy to formulate the corresponding condition on $I(X)$: If R is a ring and $I \subset R$ is an ideal, then the set

$$\operatorname{rad} I := \{f \in R | f^m \in I \quad \text{for some integer } m\}$$

is an ideal. (*Reason*: If f^m and $g^n = 0$, then $(af + bg)^{n+m} = 0$, since it is a sum of polynomials each divisible by either f^n or g^m.) It is called the **radical** of I. An ideal I is called a **radical ideal** if $I = \operatorname{rad} I$. It follows at once that R/I is a reduced ring iff I is a radical ideal. Thus, the ideals $I(X)$ are all radical ideals.

Not even every radical ideal in S can occur as $I(X)$: For example, the ideal $I = (x^2 + 1) \subset \mathbf{R}[x]$ is radical because $\mathbf{R}[x]/(x^2 + 1) \cong \mathbf{C}$ is reduced. But $Z(I) = \emptyset$, so I is not of the form $I(X)$ for any X. If k is algebraically closed, however, the situation is better. For example, every polynomial in one variable is a product of linear factors, and a polynomial $f \in k[x]$ generates a radical ideal iff it has no multiple roots. In this case if X is the set of roots of f, then $I(X) = (f)$. Hilbert's Nullstellensatz [1893] extends this to polynomial rings with many variables.

Theorem 1.6 (Nullstellensatz). *Let k be an algebraically closed field. If $I \subset k[x_1, \ldots, x_n]$ is an ideal, then*

$$I(Z(I)) = \operatorname{rad} I.$$

Thus, the correspondences $I \mapsto Z(I)$ and $X \mapsto I(X)$ induce a bijection between the collection of algebraic subsets of $\mathbf{A}^n_k = k^n$ and radical ideals of $k[x_1, \ldots, x_n]$.

We shall later give five different proofs of forms of the Nullstellensatz. The strongest and most general version is that given in Theorem 4.19. Three

more proofs are given in the exercises to Chapter 4, and the fifth is given in Chapter 13. It is worth emphasizing at the outset that the only difficult part of the Nullstellensatz as we have stated it here is the identification of ideals of the form $I(X)$ as being exactly the radical ideals; see Exercise 1.8.

We now present a sequence of remarkable consequences of the Nullstellensatz, Corollaries 1.7–1.10. (In fact, each is a statement from which the Nullstellensatz could be easily deduced.) The first gives a remarkable criterion for the solvability of a family of polynomial equations.

Corollary 1.7. *A system of polynomial equations*

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ \ldots & \\ f_m(x_1, \ldots, x_n) &= 0 \end{aligned}$$

over an algebraically closed field k has no solution in k^n iff 1 can be expressed as a linear combination

$$1 = \sum p_i f_i$$

with polynomial coefficients p_i.

Proof. By the Nullstellensatz, if $Z(f_1, \ldots, f_m) = \emptyset$, then 1 is in the radical of $(f_1, \ldots, f_m)$. The converse is obvious. □

Remark: To make this an effective criterion, it is necessary to know bounds on the degrees of the polynomials p_i that may be needed, and to know bounds on the "sizes" of their coefficients. Some bounds are known, and this is an area of active research; see, for example, Kollár [1988] and Teissier [1990].

The Nullstellensatz can be used to transfer the geometric study of algebraic varieties into algebra. First, it gives us a description of the k-algebras of the form $A(X)$.

Corollary 1.8. *If k is an algebraically closed field and A is a k-algebra, then $A = A(X)$ for some algebraic set X iff A is reduced and finitely generated as a k-algebra.*

Proof. If $A = A(X)$ for some $X \subset k^n$, then $A = k[x_1, \ldots, x_n]/I(X)$ is generated as a k-algebra by $x_1, \ldots, x_n$. Since $I(X)$ is a radical ideal, A is reduced.

Conversely, if A is a finitely generated k-algebra, then after choosing generators we may write $A = k[x_1, \ldots, x_n]/I$ for some ideal I. Since A is reduced, I is a radical ideal. Thus $I = I(Z(I))$ by the Nullstellensatz, and we may take $X = Z(I)$. □

Because of the result of Corollary 1.8, reduced finitely generated k-algebras are often called **affine k-algebras**, or, when it is not necessary to refer explicitly to the field k, simply **affine rings**.

To get X from $A(X)$ with the idea of Corollary 1.8 we must choose a set of k-algebra generators for $A(X)$. But it turns out that X is in a certain sense independent of this choice. First, note that for any field the ideal of polynomials in $k[x_1, \ldots, x_n]$ vanishing at the point $p = (a_1, \ldots, a_n) \in \mathbf{A}^n$ is $\mathfrak{m}_p := (x_1 - a_1, \ldots, x_n - a_n)$. (*Reason*: It is obvious that the given ideal is in $I(p)$; but, on the other hand, factoring out $\mathfrak{m}_p$ identifies the variables x_i with the scalars a_i, so $k[x_1, \ldots, x_n]/\mathfrak{m}_p = k$, and we see that $\mathfrak{m}_p$ is a maximal ideal.) The Nullstellensatz shows that every maximal ideal has this form.

Corollary 1.9. *Let k be an algebraically closed field and let $X \subset \mathbf{A}^n$ be an algebraic set. Every maximal ideal of $A(X)$ is of the form $\mathfrak{m}_p := (x_1 - a_1, \ldots, x_n - a_n)/I(X)$ for some $p = (a_1, \ldots, a_n) \in X$. In particular, the points of X are in one-to-one correspondence with the maximal ideals of the ring $A(X)$.*

Proof. The maximal ideals of $A(X)$ correspond to the maximal ideals of $k[x_1, \ldots, x_n]$ containing $I(X)$, so it suffices to treat the case $X = \mathbf{A}^n$, $A(X) = k[x_1, \ldots, x_n]$. To prove the first statement, note that any maximal ideal $\mathfrak{m}$—even any prime ideal —is a radical ideal, and thus $I(Z(\mathfrak{m})) = \mathfrak{m}$ by the Nullstellensatz. But if $p \in Z(\mathfrak{m})$, then $\mathfrak{m} \subset \mathfrak{m}_p$, and since $\mathfrak{m}$ is assumed maximal, $\mathfrak{m} = \mathfrak{m}_p$. The second statement follows at once. □

Given a reduced affine algebra A over an algebraically closed field, Corollary 1.8 tells us that $A = A(X)$ for some algebraic set X, and Corollary 1.9 gives us X as a set. But having X as a set is not enough; we would like to show that A reflects all the "structure of X as an algebraic set." For this we must know when two algebraic sets should be considered isomorphic or, better, what the natural maps are between algebraic sets. For simplicity, we shall assume throughout the following discussion that the ground field k is algebraically closed.

The natural maps of one algebraic set $X \subset k^n$ to another, $Y \subset k^m$ are those that are the restrictions of polynomial maps

$$F : (a_1, \ldots, a_n) \mapsto (f_1(a_1, \ldots, a_n), \ldots, f_m(a_1, \ldots, a_n))$$

from k^n to k^m, and such maps are called **morphisms** (or **polynomial maps,** or **regular maps**) from X to Y. We can use the same polynomials f_i to define a map of rings

$$F^{\#} : k[y_1, \ldots, y_m] \to k[x_1, \ldots, x_n]$$

sending y_i to $f_i(x_1, \ldots, x_n)$. To say that F restricts to a map carrying X to Y is to say that if $g \in I(Y)$, then

$$F^{\#}(g) = g(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

vanishes on X; equivalently, $F^{\#}(g) \in I(X)$. Thus $F^{\#}$ induces a map of k-algebras, which we also call $F^{\#}$,

$$F^{\#} : A(Y) = k[y_1, \ldots, y_m]/I(Y) \to k[x_1, \ldots, x_n]/I(X) = A(X).$$

If we regard $A(X)$ and $A(Y)$ as rings of functions on X and Y, then the map $F^{\#}$ is simply "composition with F." This shows that two maps F and F' with the same restriction to X induce the same map $F^{\#} : A(Y) \to A(X)$.

This process can be reversed: Given any map of k-algebras $\varphi : A(Y) \to A(X)$, we may choose representatives f_i in $k[x_1, \ldots, x_n]$ of the elements $\varphi(y_i)$ and thus get a set of m polynomials that define a map $F : k^n \to k^m$ carrying X to Y, and such that $\varphi = F^{\#}$. In terms of the description of X and Y as sets given in Corollary 1.9, the map F acts as follows: If $p \in X$, then p corresponds to a maximal ideal $\mathfrak{m}_p \subset A(X)$, and from the form of $\mathfrak{m}_p$ given in Corollary 1.9 we see that $A(X)/\mathfrak{m}_p = k$. The composite map $A(Y) \to A(X) \to A(X)/\mathfrak{m}_p = k$ is a surjection because there is a copy of k in $A(Y)$ that maps to the copy of k in $A(X)$ that already surjects to $k = A(X)/\mathfrak{m}_p$. Thus, the kernel of the composite map is a maximal ideal of $A(Y)$. By Corollary 1.9 again, this maximal ideal corresponds to a point q of Y. The map F takes p to q.

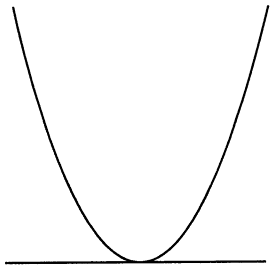
The beauty of this description of the morphisms from X to Y is that it is independent of the description of $A(X)$ and $A(Y)$ as quotients of particular polynomial rings—that is, it is independent of the particular embeddings of $X \subset k^n$ and $Y \subset k^m$! In particular, we see that X and Y are isomorphic by polynomial maps iff $A(X)$ and $A(Y)$ are isomorphic as k-algebras.

In sum, the Nullstellensatz gives us:

Corollary 1.10. *The category of affine algebraic sets and morphisms (over an algebraically closed field k) is equivalent to the category of affine k-algebras with the arrows reversed.*

Here the notion of equivalence of categories is just that the objects correspond to one another, and the morphisms do too (a formal definition may be found in Appendix A5). Using Corollary 1.10 we recapture the whole geometric picture of varieties and their maps in algebra, and at the same time we have a whole wealth of geometric ideas to bring into the study of rings, at least rings without nilpotent elements. Some further steps are outlined in Exercises 1.24 and 1.25. (Starting in the 1950s Grothendieck took the step to arbitrary rings by generalizing the geometric side of the equivalence, the affine algebraic sets, to **affine schemes**; but we shall leave this to a course on algebraic geometry. See, for example, Eisenbud and Harris [1992] and Hartshorne [1977].)

Example. The ring represented by the following figure



might be

$$k[x, y]/(y - x^2) \cap (y) = k[x, y]/(y^2 - x^2 y).$$

Of course it is hard to tell from the picture exactly which curve tangent to the given line is meant, and we have simply chosen one—the parabola tangent to the x-axis—for the purpose of writing down a definite example. This is typical: The picture captures some qualitative aspect of a ring, but generally does not specify it completely. After a little experience, this will cause the reader no trouble. Further examples for the reader to try are given in the exercises to this chapter.

1.7 Geometric Invariant Theory

Let G be a group acting on a set X. The **quotient** space X/G is by definition the set of orbits of G in X, and there is a natural **projection map** $\pi : X \to X/G$ taking each element $x \in X$ to its orbit $Gx \in X/G$. If X is a topological space, then X/G is naturally a topological space too, if we define an open set of X/G to be a set whose preimage in X is open.

If X is an algebraic variety over a field k, and G acts by polynomial maps, one might hope that X/G could be made into an algebraic variety in such a way that π is a morphism. In general, however, this is not possible. For example, if X is the affine line and G is the multiplicative group of nonzero elements in k, then G has only two orbits on X—the set of nonzero elements and the set $\{0\}$. But 0 is in the closure of the set of nonzero elements in the Zariski topology! The quotient space X/G thus consists of two points, and one of them is in the closure of the other. But any finite algebraic set has the discrete topology, where every point is closed. Thus X/G cannot be made into an algebraic variety in such a way that π is a morphism. However, if we restrict our attention to the open set of nonzero elements in X, then the quotient, consisting of just one point, is a perfectly nice affine algebraic variety. It turns out that this trivial example is rather typical.

If $A(X)$ is the affine coordinate ring of X, then G will act on $A(X)$ by composition (if $g \in G$, and $f \in A(X)$ is a polynomial function, then fg

is a polynomial function). Since the invariant functions, the elements of $A(X)^G$, are exactly those polynomial functions that are constant on the orbits, we can regard $A(X)^G$ as a ring of functions on the quotient X/G. If X/G or some large subset of it is to be an affine algebraic variety in a way compatible with the quotient map π, then $A(X)^G$ should be its coordinate ring. Geometric invariant theory, as developed by David Mumford (see Mumford and Fogarty [1982]), is the study of such quotients X/G and the algebraic varieties that "approximate" them. Here is how it begins:

Suppose that we are in a situation, such as the ones given by Corollary 1.5, where $R := A(X)^G$ is a finitely generated k-algebra. The algebra R is a subring of $A(X)$, so R is reduced. Suppose further that k is algebraically closed. By the Nullstellensatz, $R = A(Y)$ for some algebraic set Y, which may be identified with the set of maximal ideals of R. Furthermore, there is a natural map $\pi : X \to Y$ determined as follows: A point $x \in X$ corresponds to a maximal ideal $\mathfrak{m}_x$ of $A(X)$. By Corollary 1.9, the composite map $k \to A(X) \to A(X)/\mathfrak{m}_x$ is an isomorphism. It follows that the composite map $k \to R \to R/(R\cap\mathfrak{m}_x)$ is an isomorphism, so that $R\cap\mathfrak{m}_x$ is a maximal ideal of R. Let $y \in Y$ be the point corresponding to $R \cap \mathfrak{m}_x$. We set $\pi(x) = y$. As shown in the discussion preceding Corollary 1.10, π is actually a polynomial map.

In addition, we claim that π factors through the set X/G. Indeed, since R is invariant under g, we have $R \cap \mathfrak{m}_x = g(R \cap \mathfrak{m}_x) = R \cap g(\mathfrak{m}_x)$. Since $g(\mathfrak{m}_x) = \mathfrak{m}_{g^{-1}x}$, this says that $\pi(g^{-1}x) = \pi(x)$; that is, $\pi : X \to Y$ factors through a map $X/G \to Y$.

Under good circumstances, the map $X/G \to Y$ is surjective. If we are in situation where R is a summand of $A(X)$, as in the cases Hilbert treated, then for any maximal ideal $\mathfrak{n}$ of R we have $\mathfrak{n}A(X) \neq A(X)$. Thus there is a maximal ideal $\mathfrak{m}$ of $A(X)$ containing $\mathfrak{n}A(X)$ and with it $\mathfrak{n}$. Since $\mathfrak{n}$ is maximal, we must have $\mathfrak{m} \cap R = \mathfrak{n}$.

Now suppose that G is a finite group acting by linear transformations on $X = \mathbf{A}^r$, over an algebraically closed field of characteristic 0. We have seen that $A(X)^G$ is finitely generated and a summand of $A(X)$, so π induces an epimorphism $\mathbf{A}^r/G \to Y$ as above. We shall prove in Chapter 13 that the map π identifies $\mathbf{A}^r/G$ with Y, giving the quotient a natural structure of an algebraic variety. The same would be true, by a more careful analysis, even without the assumption that k is of characteristic 0.

The cases of greatest interest in invariant theory are those where $k = \mathbf{C}$ and G is a group such as $\mathrm{SL}_n(\mathbf{C})$ or $\mathrm{GL}_n(\mathbf{C})$—not a finite group. Such cases arise when one wants to make a "moduli" space, a space whose points correspond to the isomorphism classes of certain algebraic sets. (The idea is that one finds some "canonical" embedding for the algebraic sets, so that the isomorphisms between the embedded objects become the linear automorphisms of the ambient space.) Again, in these cases $A(X)^G$ is finitely generated, but X/G may not be an algebraic variety, and the set corresponding to $A(X)^G$ may be a further quotient, as in the example of the

multiplicative group acting on the affine line. This delicate and important phenomenon is the subject of geometric invariant theory.

1.8 Projective Varieties

Kepler in 1604 and Desargues in his book in 1639 realized that the introduction of imaginary points could substantially simplify Euclidean geometry. Each line in the plane was given one new point, "at infinity," in such a way that two parallel lines would meet at their points at infinity. In this way many geometric results became simpler (for example, the statement that *every* two distinct lines meet in exactly one point), and a remarkable duality between points and lines was introduced. For example, the parenthetical statement in the last sentence is dual to the statement that through *every* pair of distinct points—including points at infinity—passes exactly one line. Of course, this makes unavoidable the idea that the set of points at infinity form a line.

Although this development came at almost the same time as the introduction of coordinates in geometry by Fermat and Descartes (Descartes' book was published in 1637), it was nearly 200 years later, in the works of Möbius (1827) and especially Plücker (1830), that the plane with these additional points was coordinatized. Plücker's system for coordinatization is the one in use today. By means of it we can define algebraic sets in the projective plane. It turns out that the Nullstellensatz can be used to make these correspond to homogeneous ideals in the polynomial ring in three variables. We turn now to this correspondence.

If k is a field, then the **projective r-space over k**, written $\mathbf{P}^r(k)$ or simply $\mathbf{P}^r$, is the set of one-dimensional subspaces (meaning lines through the origin) of an $(r+1)$-dimensional vector space over k. A one-dimensional subspace $L \subset k^{r+1}$ may be represented by a point $(a_0, \ldots, a_r) \neq (0, \ldots, 0)$ of L, the representation being unique up to a nonzero scalar multiple. The elements $a_0, \ldots, a_r$ are called **homogeneous coordinates** of the point L. Via this representation, $\mathbf{P}^r(k)$ may be regarded as the set of $(r+1)$-tuples $(a_0, \ldots, a_r)$ of elements of k, modulo the equivalence relation $(a_0, \ldots, a_r) \sim (ba_0, \ldots, ba_r)$ for $b \neq 0$ in k.

Given a polynomial in $r+1$ variables $f(x_0, \ldots, x_r)$, and a point L represented by an $(r+1)$-tuple $(a_0, \ldots, a_r)$, it makes no sense to "evaluate f at L," because the value $f(a_0, \ldots, a_r)$ depends on the representative chosen. But if f is a homogeneous polynomial of degree d, then for $b \in k$ we have

$$f(ba_0, \ldots, ba_r) = b^d f(a_0, \ldots, a_r),$$

so the statement that $f(a_0, \ldots, a_r) = 0$ is independent of the representative, and it makes sense to say whether or not f vanishes at L.

Let $S = k[x_0, \ldots, x_r]$, and let S_d be the vector space of all forms of degree d, so that we have $k = S_0$, and

$$S = S_0 \oplus S_1 \oplus S_2 \oplus \cdots .$$

Since $S_i S_j \subset S_{i+j}$ for $i, j \geq 0$, we may regard S with this decomposition as a graded ring, graded "by degree." Given any homogeneous ideal I of S, we define the **projective algebraic set $Z(I)$ associated to I** to be

$$Z(I) = \{(a_0, \ldots, a_r) \in \mathbf{P}^r(k) | f(a_0, \ldots, a_r) = 0 \\ \text{for all homogeneous } f \in I\}.$$

The irrelevant ideal corresponds to the empty set—whence its name.

Again, there is a sort of inverse operation: Given any subset $X \subset \mathbf{P}^r(k)$, we define $I(X)$ to be the homogeneous ideal in S generated by all forms vanishing on X. If k is algebraically closed, then, just as in the affine case, the Nullstellensatz gives a bijection between the set of radical homogeneous ideals of S other than the irrelevant ideal and the set of projective algebraic sets in $\mathbf{P}^r(k)$; the only additional observation required is that the radical of a homogeneous ideal is homogeneous. (*Reason*: Suppose I is a homogeneous ideal, and $f = f_d + f_{d+1} + \cdots + f_e \in \operatorname{rad} I$, where each f_c is a homogeneous form of degree c. If $f^n \in I$, then since I is homogeneous, the homogeneous components of f^n are in I too. The lowest degree component is f_d^n. Thus $f_d \in \operatorname{rad} I$. Subtracting f_d from f and repeating the argument, we see that each homogeneous component of f is in rad I, so rad I is homogeneous). Note that if we included the irrelevant ideal, we would not get a one-to-one correspondence: Both the irrelevant ideal and the "unit ideal" $(1) = S$ would correspond to the empty set.

The graded ring $S/I(X)$ is called the **homogeneous coordinate ring of X**. It is an invariant not of X alone (as in the affine case), but of X together with its embedding into projective space.

One way to view projective algebraic sets is as conical algebraic sets in affine $(r+1)$-space—that is, sets Y such that $(a_0, \ldots, a_r) \in Y$ implies $(aa_0, \ldots, aa_r) \in Y$ for all scalars $a \in k$. It is not hard to show, conversely, that if k is infinite then the ideal of any such cone is a homogeneous ideal (this shows that the correspondence between projective algebraic sets and homogeneous ideals may be regarded as a special case of the Nullstellensatz in Theorem 1.6 rather than a parallel theorem).

Projective space may be viewed as affine space "completed" by adding some "points at infinity" (in case the ground field k is $\mathbf{C}$, we can take "completed" to mean "compactified"). We now describe this view.

Consider the complement U in $\mathbf{P}^r$ of the hyperplane H defined by the equation $x_0 = 0$, that is, $U = \{(a_0, \ldots, a_r) \in \mathbf{P}^r | a_0 \neq 0\}$. Since the coordinates are defined up to multiplication by a nonzero scalar, every point $(a_0, \ldots, a_r) \in U$ can be represented uniquely in the form $(1, b_1, \ldots, b_r)$, with $b_i = a_i/a_0$. The association $(a_0, \ldots, a_r) \mapsto (b_1, \ldots, b_r)$ is a bijection

between U and $\mathbf{A}^r$, so we have expressed $\mathbf{P}^r$ as a union: $\mathbf{P}^r = U \cup H = \mathbf{A}^r \cup H$. We call H the **hyperplane at infinity**. Note that H may be identified with $\mathbf{P}^{r-1}$, so we may continue this decomposition and obtain $\mathbf{P}^r = \mathbf{A}^r \amalg \mathbf{A}^{r-1} \amalg \cdots \amalg \mathbf{A}^0$, where we have written $\amalg$ for disjoint union. Alternately, since we could have started with any variable x_i in place of x_0, we have defined a covering of $\mathbf{P}^r$ by copies of $\mathbf{A}^r$ (not disjoint), each the complement of one of the hyperplanes $x_i = 0$.

For these identifications to be useful, we must see that an algebraic set in $\mathbf{P}^r$ meets U in an affine algebraic set. To this end note that if $X \subset \mathbf{P}^r$ is an algebraic set defined by homogeneous polynomial equations $F_i(x_0, \ldots, x_r) = 0$, then $X \cap U$ may be described by the polynomial equations $f_i(x_1, \ldots, x_r) = F_i(1, x_1, \ldots, x_r) = 0$; thus $X \cap U$ is naturally an algebraic set in $\mathbf{A}^r$.

Every polynomial $f(x_1, \ldots, x_r)$ may be written in the form $F(1, x_1, \ldots, x_r)$ for some homogeneous polynomial $F(x_0, \ldots, x_r)$. For example, let d be the degree of f, and let F be the result of multiplying each homogeneous component of f by a power of x_0 to bring up its degree to d. More formally, we may write

$$F(x_0, \ldots, x_r) = x_0^d f(x_1/x_0, \ldots, x_r/x_0).$$

It follows that $F(1, x_1, \ldots, x_r) = f(x_1, \ldots, x_r)$. The form F is called the **homogenization of f with homogenizing variable x_0**. The existence of such homogenizations shows that every algebraic set in $\mathbf{A}^r$ is the intersection of U with an algebraic set in $\mathbf{P}^r$.

These remarks show that it is reasonable to identify U with $\mathbf{A}^r$. They also suggest a natural operation: Given an affine algebraic set $X \subset \mathbf{A}^r$, define the **projective closure** $\bar{X}$ of X in $\mathbf{P}^r$ to be the smallest algebraic set intersecting $U = \mathbf{A}^r$ in X. The homogeneous ideal of the set $\bar{X}$ is generated by the homogeneous forms $F(x_0, \ldots, x_r)$ such that $f(x_1, \ldots, x_r) := F(1, x_1, \ldots, x_r) \in I(X)$. Since $F(x_0, \ldots, x_r)$ is a power of x_0 times the homogenization of f, it follows that $I(\bar{X})$ is generated by the homogenizations of all the elements of $I(X)$. Some caution must be exercised here: It is not enough to take the homogenizations of a set of generators of $I(X)$. See Exercise 1.17 for an example.

1.9 Hilbert Functions and Polynomials

It is interesting to look for numerical invariants of a projective algebraic set $X = Z(I) \subset \mathbf{P}_k^r$. The next major result of Hilbert concerns a particularly simple and important class of such invariants, given by the dimensions of the spaces of forms I_d of degree d vanishing on X for various d. Since the space of all forms of degree d has known dimension $\binom{r+d}{r}$, knowing the dimension of I_d is equivalent to knowing the dimension of the degree d

part of the homogeneous coordinate ring $k[x_0, \ldots, x_r]/I$. Hilbert's original motivation for studying these numbers came again from invariant theory: Given the action of a group on the linear forms of a polynomial ring, he wanted to understand how the dimension of the space of invariant forms of degree d can vary with d.

The natural context is that of graded modules:

Definition. *If $R = R_0 \oplus R_1 \oplus \cdots$ is a graded ring, then a **graded module** over R is a module M with a decomposition*

$$M = \bigoplus_{-\infty}^{\infty} M_i \quad \textit{as abelian groups}$$

such that $R_i M_j \subset M_{i+j}$ for all i, j.

Definition. *Let M be a finitely generated graded module over $k[x_1, \ldots, x_r]$, with grading by degree, as in the preceding definition. The numerical function*

$$H_M(s) := \dim_k M_s$$

*is called the **Hilbert function of M**. (These dimensions are all finite; if M_s were not finite dimensional, then the submodule $\oplus_s^\infty M_i$ would not be finitely generated, contradicting Proposition* 1.4.)

Hilbert's insight was that all the information encoded in the infinitely many values of the function H_M can be read off from just finitely many of its values, and in a simple way:

Theorem 1.11 (Hilbert). *If M is a finitely generated graded module over $k[x_1, \ldots, x_r]$, then $H_M(s)$ agrees, for large s, with a polynomial of degree $\leq r - 1$.*

Definition. *This polynomial, denoted $P_M(s)$, is called the **Hilbert polynomial of M**.*

Before proving the theorem, we need a notation to indicate that we have altered a graded module M by "shifting" its grading d steps. We define $M(d)$ to be this graded module; more formally, $M(d)$ is isomorphic to M as a module and has grading defined by

$$M(d)_e = M_{d+e}.$$

$M(d)$ is sometimes referred to as the **dth twist of M**. Many natural maps of graded modules take the grading of one to the grading of the other with a shift of degrees. Using our notation, we can write them as maps of degree 0 (so that they take homogeneous elements to homogeneous elements of

the same degree) between one of the modules and a shift of the other. This makes it is easy to keep track of graded components. For example, multiplication by a linear form on a module M as above raises the degrees by 1. Thus it can be thought of as a map of degree 0 from $M(-1)$ to M. We shall use this idea in the following proof. We shall also use an elementary result about integer-valued functions.

Lemma 1.12. *Let $H(s) \in \mathbf{Z}$ be defined for all natural numbers s. If the "first difference" $H'(s) = H(s) - H(s-1)$ agrees with a polynomial of degree $\leq n-1$ having rational coefficients for $s \geq s_0$, then $H(s)$ agrees with a polynomial of degree $\leq n$ having rational coefficients for all $s \geq s_0$.*

Proof. Suppose that $Q(s)$ is a polynomial of degree $\leq n-1$ with rational coefficients such that $H'(s) = Q(s)$ for $s \geq s_0$. For any integer s set $P(s) = H(s_0)+\sum_{t=s_0+1}^{s} Q(t)$, where the sum is taken over all integers between s_0+1 and s whether $s \geq s_0+1$ or $s \leq s_0+1$. For $s \geq s_0$ we have $P(s) = H(s)$. For all s we have $P(s) - P(s-1) = Q(s)$. It follows that $P(s)$ is a polynomial of degree $\leq n$ with rational coefficients; see Exercise 1.21a and its hint for a quick proof. □

Proof of Theorem 1.11. We do induction on r, the number of variables. If $r = 0$, then M is simply a finite-dimensional graded vector space. In this case $H_M(s) = 0$ for all large s, and this is a polynomial of degree -1.

In the general case, if we let $K \subset M$ be the kernel of multiplication by x_r, we get an exact sequence of graded vector spaces, with maps of degree 0 :

$$0 \to K(-1) \to M(-1) \xrightarrow{x_r} M \to M/x_rM \to 0.$$

Taking the component of degree s of each term in this exact sequence, we see that

$$H_M(s) - H_M(s-1) = H_{M/x_rM}(s) - H_K(s-1).$$

Now both K and M/x_rM are finitely generated modules over $k[x_1, \ldots, x_{r-1}]$. By induction, the terms on the right-hand side agree for large s with polynomials of degree less than or equal to $r-2$, and we are done by Lemma 1.12. □

The Hilbert function actually includes *all* the invariants of modules that are additive in a certain sense, and the Hilbert polynomial includes all additive invariants that vanish on modules of finite length. See Exercises 19.15 and 19.16.

In the case of greatest interest, M is the homogeneous coordinate ring of a projective algebra set $X \subset \mathbf{P}^{r-1}$. Here the Hilbert function is a rich source of discrete invariants of X and its embedding. We shall see that the degree d of the Hilbert polynomial $P(s)$ is the **dimension** of X in a suitable sense, and the initial coefficient of $P(s)$, multiplied by d factorial,

is what is called the **degree** of X—the number of points in which X meets a general plane of complementary dimension in $\mathbf{P}^{r-1}$. See Exercise 1.18 for an illustration.

One may well wonder what the values $P_M(s)$ are for small values of s when they are not the values of $H_M(s)$. We shall provide an answer in terms of free resolutions. There is a different answer in terms of cohomology groups: The function that is really a polynomial is an Euler characteristic, made from the Hilbert functions of the module and all its cohomology groups. The cohomology groups have geometric interpretations, and this expression is quite useful in geometric applications.

We mention in passing two further geometric contexts in which the Hilbert polynomial appears: First, the **Riemann-Roch theorem** is a computation of the Hilbert polynomial (for a certain class of modules) that plays an enormously important role in algebraic geometry. Second, a graded module over a polynomial ring corresponds to a "coherent sheaf" on a projective space. The information contained in the coefficients of the Hilbert polynomial is usually presented in algebraic geometry by giving the **Chern classes** of this sheaf—a different set of integers, which can be deduced from the coefficients, and from which the coefficients can also be deduced. See Exercise 19.18.

1.10 Free Resolutions and the Syzygy Theorem

> The members of any group of functions, more than two in number, whose nullity is implied in the relation of double contact ... must be in syzygy.
>
> —J.J. Sylvester, 1850 (First mathematical use of the term syzygy, according to the Oxford English Dictionary)

The word syzygy, (from the Greek word for pairing (or copulation)) has long been used in English as an astronomical term for conjunctions of planets. But since the middle of the last century, and in particular since the work of Hilbert at the end of the century, its meaning has had to do with the solutions to a system of homogeneous linear equations over a ring.

The proof we have given for Theorem 1.11 is quite different from the one Hilbert gave. In place of our induction, he used free resolutions. We shall now sketch his ideas, postponing proofs until Chapters 15 and 19.

If R is a graded ring, then we shall define a **graded free R-module** to be a direct sum of modules of the form $R(d)$, for various d. Note that the nice mnemonic definition of $M(d)$ by the formula $M(d)_e = M_{d+e}$ has the at first rather annoying consequence that $R(d)_{-d} = R_0$, so $R(d)$ has its generator in degree $-d$, not d.

Definition. *A **complex of R-modules** is a sequence of modules F_i and maps $F_i \to F_{i-1}$ such that the compositions $F_{i+1} \to F_i \to F_{i-1}$ are all zero. The **homology** of this complex at F_i is the module*

$$\ker(F_i \to F_{i-1})/\operatorname{im}(F_{i+1} \to F_i).$$

*A **free resolution** of an R-module M is a complex*

$$\mathcal{F}: \ \cdots \to F_n \xrightarrow{\varphi_n} \cdots \to F_1 \xrightarrow{\varphi_1} F_0$$

of free R-modules such that coker $\varphi_1 = M$ and $\mathcal{F}$ is exact (sometimes we add "$\to 0$" to the right of $\mathcal{F}$ and then insist that $\mathcal{F}$ be exact except at F_0). We shall sometimes abuse this notation and say that an exact sequence

$$\mathcal{F}: \ \cdots \to F_n \xrightarrow{\varphi_n} \cdots \to F_1 \xrightarrow{\varphi_1} F_0 \to M \to 0$$

*is a resolution of M. The image of the map φ_i is called the ith syzygy module of M. A resolution $\mathcal{F}$ is a **graded free resolution** if R is a graded ring, the F_i are graded free modules, and the maps are homogeneous maps of degree 0. Of course only graded modules can have graded free resolutions. If for some $n < \infty$ we have $F_{n+1} = 0$, but $F_i \neq 0$ for $0 \leq i \leq n$, then we shall say that $\mathcal{F}$ is a **finite resolution of length** n.*

It is easy to see that every module has a free resolution and, if R is graded, that every graded module has a graded free resolution. To construct one, begin by taking a set of generators for M and map a free module onto M sending the free generators of the free module to the given generators of M. Let M_1 be the kernel of this map, and repeat the procedure, now starting with M_1.

Exercises 1.22 and 1.23 give two special cases in which free resolutions are not difficult to compute "by hand"; we shall eventually give far-reaching generalizations of both.

We can finally state, in the following theorem, the last of the four great results on commutative algebra in Hilbert's papers.

Theorem 1.13 (Hilbert syzygy theorem). *If $R = k[x_1, \ldots, x_r]$, then every finitely generated graded R-module has a finite graded free resolution of length $\leq r$, by finitely generated free modules.*

We shall give a constructive proof of the syzygy theorem in Chapter 15, and a different, nonconstructive proof in Chapter 19. Here we apply it to its original purpose:

Hilbert's Proof of Theorem 1.11. Let $R = k[x_1, \ldots, x_r]$. If $M = R(d)$ for some d, then

$$H_{R(d)}(s) = H_R(s+d) = \binom{s+d+r-1}{r-1},$$

which agrees for $s \geq -(d+r-1)$ with the polynomial

$$\begin{aligned} Q(s) &= (1/(r-1)!)[s+(d+r-1)] \cdot [s+(d+r-2)] \cdot \cdots \cdot [s+d] \\ &= s^{r-1}/(r-1)! + (\text{lower order terms}). \end{aligned}$$

If F is a finitely generated graded free module, then F is a direct sum of various $R(d)$, so $H_F(s)$ is a finite sum of functions of the form $H_{R(d)}(s)$.

The Syzygy theorem shows that any finitely generated graded module over $R = k[x_1, \ldots, x_r]$ has a finite graded free resolution $\mathcal{F}$

$$\mathcal{F}:\ 0 \to F_r \to \cdots \to F_0 \to M \to 0.$$

Thus

$$H_M(s) = \sum (-1)^i H_{F_i}(s)$$

is a linear combination of functions that are eventually equal to polynomials of degree $\leq r-1$. □

Note that this proof yields a simple computation for the value of $H_M(s)$ for every s, not just for large s. It also shows that the deviation from being a polynomial comes about because $R_d = 0$ for all $d < 0$—of course no nonzero polynomial has this property.

In Chapter 15 we shall see that free resolutions can be computed effectively, and this yields an effective computation of H_M and P_M (in the same chapter we shall give a much simpler method of computation).

1.11 Exercises

Noetherian Rings and Modules

Exercise 1.1:* Prove that the following conditions on a module M over a commutative ring R are equivalent (the fourth is Hilbert's original formulation; the first and third are the ones most often used). The case $M = R$ is the case of ideals.

1. M is Noetherian (that is, every submodule of M is finitely generated).
2. Every ascending chain of submodules of M terminates ("ascending chain condition").
3. Every set of submodules of M contains elements maximal under inclusion.
4. Given any sequence of elements $f_1, f_2, \ldots \in M$, there is a number m such that for each $n > m$ there is an expression $f_n = \sum_{i=1}^{m} a_i f_i$ with $a_i \in R$.

Exercise 1.2 (Emmy Noether): Prove that if R is Noetherian, and $I \subset R$ is an ideal, then among the primes of R containing I there are only finitely many that are minimal with respect to inclusion (these are usually called the **minimal primes of *I***, or the **primes minimal over *I***) as follows: Assuming that the proposition fails, the Noetherian hypothesis guarantees the existence of an ideal I maximal among ideals in R for which it fails. Show that I cannot be prime, so we can find elements f and g in R, not in I, such that $fg \in I$. Now show that every prime minimal over I is minimal over one of the larger ideals (I, f) and (I, g).

With Hilbert's basis theorem and the Nullstellensatz (see Exercise 1.9), Exercise 1.2 gives one of the fundamental finiteness theorems of algebraic geometry: An algebraic set can have only finitely many irreducible components. Originally the result was proved by difficult inductive arguments and elimination theory. For a further discussion of the significance of this result see the beginning of Chapter 3, and particularly example 2 there. The result of this exercise is strengthened in Theorem 3.1.

Exercise 1.3: Let M' be a submodule of M. Show that M is Noetherian iff both M' and M/M' are Noetherian.

An Analysis of Hilbert's Finiteness Argument

Exercise 1.4:* We have seen from Corollary 1.3 that any finitely generated algebra over a field is Noetherian. The converse is quite false, and we shall see many important examples of rings that are Noetherian but not finitely generated (for instance localizations and completions). But the converse is true for graded rings R where R_0 is a field, as the following result shows.

Let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded ring. Prove that the following are equivalent:

1. R is Noetherian.
2. R_0 is Noetherian and the irrelevant ideal $R_1 \oplus R_2 \oplus \cdots$ is finitely generated.
3. R_0 is Noetherian and R is a finitely generated R_0-algebra.

Exercise 1.5:* Although the Noetherian property does not usually pass from a ring to a subring, it does when the subring is a summand:

Let $R \subset S$ be rings, and assume that R is a summand of S as an R-module, that is, there is a homomorphism $\varphi : S \to R$ of R-modules fixing every element of R. Prove that if S is Noetherian, then R is Noetherian.

Some Rings of Invariants

Exercise 1.6: The following proof of the assertions of Example 1.1 is from Van der Waerden [1971]. We shall systematically develop this method in Chapter 15. Let $\sum$ and $f_1, \ldots, f_r$ be as defined in Example 1.1.

Order the monomials of the polynomial ring $S = k[x_1, \ldots, x_r]$ according to the **degree-lexicographic** order, defined as follows: Let $A = x_1^{m_1} \cdots x_r^{m_r}$ and $B = x_1^{n_1} \cdots x_r^{n_r}$ be two monomials. We say that $A > B$ if either $\deg A > \deg B$, or else $\deg A = \deg B$ and the sequence of exponents $(m_1, \ldots, m_r)$ is greater than the sequence $(n_1, \ldots, n_r)$ in the lexicographic order; that is, the difference $m_i - n_i > 0$ for the first index i for which it is not zero.

Given any polynomial $p(x_1, \ldots, x_r)$, we define the **initial term** of p to be the term involving the greatest monomial in the order $>$.

a. Show that for each monomial A there are only finitely many monomials B such that $A > B$.

b.* Show that if p is invariant under $\sum$, then the initial term of p is an element of k times a monomial $x_1^{m_1} \cdots x_r^{m_r}$ with $m_1 \geq m_2 \geq \cdots \geq m_r$.

c. Show that the initial term of the product $f_1^{\mu_1} \cdots f_r^{\mu_r}$ is $x_1^{m_1} \cdots x_r^{m_r}$, where $m_i = \sum_{j \geq i}^r \mu_j$.

d. Show that the function $\mathbf{Z}^r \to \mathbf{Z}^r$ defined by

$$(\mu_1, \ldots, \mu_r) \mapsto (m_1, \ldots, m_r) \text{ with } m_i = \sum_{j \geq i}^{r} \mu_j$$

is a monomorphism. Conclude that a monomial $x_1^{m_1} \cdots x_r^{m_r}$ with $m_1 \geq m_2 \geq \cdots \geq m_r$ is the initial monomial of a unique product of f_i.

e.* Now show that any element of S^Σ can be written uniquely as a polynomial in the f_i.

Exercise 1.7:* a. (The simplest group action whose ring of invariants is not a polynomial ring) Suppose that k is a field of characteristic $\neq 2$. Let the generator g of the group $G := Z/2$ act on the polynomial ring $k[x, y]$ in two variables by sending x to $-x$ and y to $-y$. Show that the ring of invariants is $k[x^2, xy, y^2]$. Prove that $k[x^2, xy, y^2] \cong k[u, v, w]/(uw - v^2)$. Show that this is not isomorphic to any polynomial ring over a field. (A theorem of Shepard, Todd and Chevalley shows that if a finite group acts by linear transformations of the variables on a polynomial ring, then the ring of invariants is isomorphic to a polynomial ring iff the group is generated by "pseudo-reflections," where an element is a **pseudo-reflection** if it acts as the identity on a hyperplane. See Sturmfels [1992], Section 2.4.)

b. More generally, let G be any finite abelian group, acting linearly on the space of linear forms of the ring $S = k[x_1, \ldots, x_r]$. Assume that G acts by characters; that is, assume that there are homomorphisms $\alpha_i : G \to k^\times$, and $g(x_i) = \alpha_i(g)x_i$ for all $g \in G$, where $k^\times$ is the multiplicative group of the field k. (As long as the characteristic of k does not divide the order of G, this could be achieved, by a suitable choice of variables, for any action of G.) Show that the invariants of G are generated by those monomials $\Pi x_i^{a_i}$ whose exponent vectors $(a_1, \ldots, a_r)$ are in the kernel of a map from $\mathbf{Z}^r$ to a certain

finite abelian group. Conclude that the quotient field of S^G is isomorphic to a field of rational functions in r variables. (Emmy Noether asked, in a famous paper, whether the last statement is true for finite nonabelian groups as well. It is not; see, for example, Saltman [1982] for a survey.)

Algebra and Geometry

Exercise 1.8 (A formal Nullstellensatz): Let $\mathfrak{X}$ and $\mathfrak{J}$ be partially ordered sets, and suppose that $I : \mathfrak{X} \to \mathfrak{J}$ and $Z : \mathfrak{J} \to \mathfrak{X}$ are functions such that

i) I and Z reverse the order in the sense that $x \leq y \in \mathfrak{X}$ implies $I(x) \geq I(y)$, and $i \leq j \in \mathfrak{J}$ implies $Z(i) \geq Z(j)$.

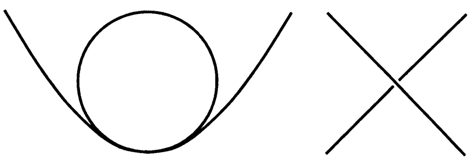
ii) ZI and IZ are increasing functions, in the sense that $x \in \mathfrak{X}$ imples $ZI(x) \geq x$, and $i \in \mathfrak{J}$ imples $IZ(i) \geq i$.

a. Show that I and Z establish a one-to-one correspondence between the subsets $I(\mathfrak{X}) \subset \mathfrak{J}$ and $Z(\mathfrak{J}) \subset \mathfrak{X}$.

b. Let k be a field. Call an ideal $I \subset k[x_1, \ldots, x_n]$ **formally radical** if it is of the form $Z(X)$ for some set $X \subset k^n$. Use part a to prove that there is a one-to-one correspondence between formally radical ideals and algebraic subsets of k^n. (Hilbert's Nullstellensatz identifies the formally radical ideals with the ordinary radical ideals when k is algebraically closed.)

Exercise 1.9: Let $S = k[x_1, \ldots, x_r]$, with k an algebraically closed field. Show that under the correspondence of radical ideals in S and algebraic subsets of $\mathbf{A}^r$, the prime ideals correspond to the algebraic sets that cannot be written as a proper union of smaller algebraic sets.

Exercise 1.10: Find rings to represent the following figures.

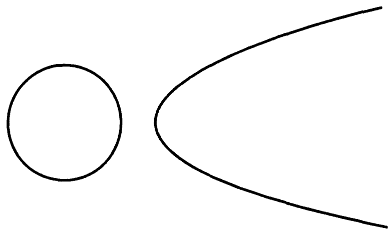


The first represents the union of a circle and a parabola in the plane, and the second shows the union of two skew lines in 3-space. (You may use the Nullstellensatz to prove that your answer is right.)

Exercise 1.11:* When we draw pictures representing algebraic sets, we often draw the same picture for all ground fields k, although by rights it generally represents best the case $k = \mathbf{R}$. In fact, we are usually interested

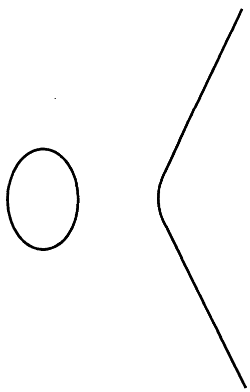
in the case of an algebraically closed field k, such as $k = \mathbf{C}$, where the Nullstellensatz applies. The main way in which the pictures can be misleading is illustrated by the following examples.

a. If $k = \mathbf{R}$ then the ring $k[x, y]/(x^2 + y^2 - 1) \cap (y - 2 - x^2)$ corresponds to the union of a circle and a parabola.

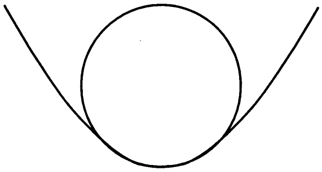


If $k = \mathbf{C}$, show that there are four points in the intersection of these two components. Show that there is a bijection given by polynomial maps between the parabola and the line $x = 0$. Show further that "projection from the north pole" gives a bijection (given by rational functions) between the circle minus one point and the line minus two points. (Can you find such a bijection between the circle and the line minus one point?)

b. Show that the polynomial $f(x, y) = y^2 - (x - 1)x(x + 1)$ is irreducible over any field k. Thus $X = Z(f) \subset \mathbf{A}^2$ is an irreducible algebraic set. This is not so obvious from the real picture, which approximately resembles the following image.

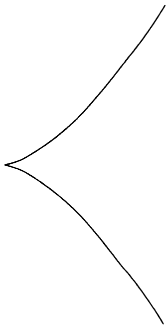


Exercise 1.12: Find equations for a parabola meeting a circle just once in the complex plane, represented by the following picture:



Exercise 1.13: Suppose that I is an ideal in a commutative ring. Show that if $\operatorname{rad} I$ is finitely generated, then for some integer N we have $(\operatorname{rad} I)^N \subset I$. Conclude that in a Noetherian ring the ideals I and J have the same radical iff there is some integer N such that $I^N \subset J$ and $J^N \subset I$. Use the Nullstellensatz to deduce that if $I, J \subset S = k[x_1, \ldots, x_r]$ are ideals and k is algebraically closed, then $Z(I) = Z(J)$ iff $I^N \subset J$ and $J^N \subset I$ for some N.

Exercise 1.14:* Not all interesting graded rings are generated by forms of degree 1, as are the homogeneous coordinate rings of projective varieties. For example, $k[x,y]/(y^2 - x^3)$, the ring corresponding to the cusp becomes



a graded ring if we give x degree 2 and y degree 3. Prove that the map $k[x,y] \to k[t]$ sending x to t^2 and y to t^3 induces an isomorphism

$$k[x,y]/(y^2 - x^3) \cong k[t^2, t^3] \subset k[t].$$

Graded Rings and Projective Geometry

Exercise 1.15 (Classification of conics and quadrics): Here is an example of the simplification brought to geometry by the idea of projective space. As the reader probably remembers from high school, a conic in the affine real plane $\mathbf{R}^2$ (that is, the locus defined by a quadratic equation in two variables x and y with real coefficients) belongs to one of the following eight types:

a. The empty set (as with $x^2 + y^2 + 1 = 0$)

b. A single point (as with $x^2 + y^2 = 0$)

c. A line ($x^2 = 0$)

d. The union of two coincident lines ($xy = 0$)

e. The union of two parallel lines ($x(x-1) = 0$)

f. A parabola ($y - x^2 = 0$)

g. A hyperbola ($xy - 1 = 0$)

h. An ellipse ($x^2 + 2y^2 - 1 = 0$)

Any two examples of one of these types differ only by an invertible linear transformation of the coordinates.

a. Show that in the complex affine plane $\mathbf{C}^2$ there are only five types of loci defined by equations of degree 2: Types a and b disappear, and types g and h coincide.

b.* Show that in the complex projective plane $\mathbf{P}^2(\mathbf{C})$ there are only three types of loci represented by quadratic equations; they are represented by types c, d, and h on the above list. More generally, there are exactly n types of nonzero quadratic forms in n variables, classified by rank (where the rank of a quadratic form $\sum_{i<j} a_{ij}x_ix_j$ is defined to be the rank of the symmetric matrix (a_{ij})).

c. Show that the different types in part (a) correspond to the relative placement of the conic and the line at infinity, in the sense that a parabola is a rank-3 conic tangent to the line at infinity, while an ellipse/hyperbola is a rank-3 conic meeting the line at infinity at two distinct points. The classification over the real numbers may be recovered from the position of these points: A real rank-3 conic meeting the line at infinity in two points is a hyperbola if the points are real, and an ellipse if the points are nonreal (they are then conjugate complex points). If the affine plane is represented by points $(x, y, z) \in \mathbf{P}^2$ with $z = 1$, the ellipse is a circle iff it meets the line at infinity in the points $(1, i, 0)$ and $(1, -i, 0)$, the "circular points at infinity."

Exercise 1.16: a. Let I be a homogeneous ideal in $S = k[x_0, \ldots, x_r]$, and suppose that the projective algebraic set corresponding to I is nonempty. Let $Y \subset k^{r+1}$ be the affine algebraic set associated to I. Show that Y is a union of one-dimensional subspaces of k^{r+1}, and that these one-dimensional subspaces are precisely the points of the projective algebraic set associated to I.

b.* Show also that if k is an infinite field and $X \subset \mathbf{A}^r(k)$ is a union of lines through the origin, then $I(X)$ is a homogeneous ideal.

Exercise 1.17:* Let $I \subset k[x_1, x_2, x_3]$ be the ideal $(x_1^2 + x_2, x_1^2 + x_3)$, and let $X \subset \mathbf{A}^3$ be the affine algebraic set $Z(I)$. Let $\bar{X} \subset \mathbf{P}^3$ be the projective

closure of X. Show that the homogeneous ideal $I(\bar{X})$ is not generated by the homogenizations of $x_1^2+x_2$ and $x_1^2+x_3$. We shall return to this subject in Chapter 15.

Hilbert Functions

Exercise 1.18: Let k be a field. Compute the Hilbert function and polynomial for the ring

$$k[x,y,z,w]/(x,y)\cap(z,w)$$

corresponding to the disjoint union of two lines in projective 3-space. Compare these to the Hilbert function and polynomial of the ring corresponding to one projective line, $k[x,y]$.

Exercise 1.19: Let k be a field. Let $I \subset k[x,y,z,w]$ be the ideal generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x & y & z \\ y & z & w \end{pmatrix},$$

that is, $I = (yw - z^2, xw - yz, xz - y^2)$.
Show that $R = k[x,y,z,w]/I$ is a finitely generated free module over $S = k[x,w]$. Exhibit a basis for R as an S-module. Show that there is a ring homomorphism $R \to k[s,t]$ such that $x \mapsto s^3$, $y \mapsto s^2t$, $z \mapsto st^2$, $w \mapsto t^3$. Use the basis you constructed to show that it is a monomorphism. Conclude that I is prime. From the rank of R as a free S-module, and the degrees of the generators, deduce the Hilbert function of R. Show that R is not finitely generated as a module over $k[x,y]$.

Exercise 1.20:* Given a number s_0, find an example of a graded $k[x_1,\ldots,x_r]$-module M generated by elements of degree 0 for which the function $H_M(s)$ is not equal to the Hilbert polynomial $P_M(s)$ for any $s < s_0$. If you find this too easy, can you find torsion-free $k[x_1,\ldots,x_r]$-modules of this sort?

Exercise 1.21: Consider the subring $T \subset \mathbf{Q}[n]$ of rational polynomials that take integral values at sufficiently large integers. T is of interest to us because it contains all the Hilbert polynomials discussed in this chapter. The ring T obviously contains $\mathbf{Z}[n]$, but it is larger: It contains things like $\binom{n}{2} = (n^2-n)/2$.

a.* Let $F(n)$ be a function defined for sufficiently large integers n, and set $G(n) := F(n+1) - F(n)$. Show that $F(n) \in \mathbf{Q}[n]$ iff $G(n) \in \mathbf{Q}[n]$, and that if these conditions are satisfied then $\deg F = 1 + \deg G$.

b. Show by induction on the degree that T is a free abelian group with basis given by the functions

$$F_k = \binom{n}{k} = n(n-1)\cdots(n-k+1)/k! \qquad 0 \le k \le \infty,$$

where F_k is a polynomial function in n of degree k.

c. Although $\mathbf{Q} \otimes_{\mathbf{Z}} T = \mathbf{Q}[n]$, the ring T itself is not finitely generated as an algebra over $\mathbf{Z}$; show that T is not even Noetherian. We shall meet T again as a free divided power algebra in Appendix A2.

Free Resolutions

Exercise 1.22: Let $R = k[x]$. Use the structure theorem for finitely generated modules over a principal ideal domain to show that every finitely generated R-module has a finite free resolution.

Exercise 1.23: Let $R = k[x]/(x^n)$. Compute a free resolution of the R-module $R/(x^m)$, for any $m \le n$. Show that the only R-modules with finite free resolutions are the free modules.

Spec, max-Spec, and the Zariski Topology

An ideal $I \subset R$ is called a **prime** ideal if R/I is an integral domain. I is a **maximal** ideal if R/I is a field, so maximal ideals are prime. The set of all prime ideals of a ring R is called the **spectrum** of R, written **Spec** R, and the set of all maximal ideals is usually denoted by the typographically awkward but reasonably descriptive name **max-Spec** R.

Exercise 1.24: In the text we defined the Zariski topology on an algebraic set over any algebraically closed field k. We may identify X with the set max-Spec $A(X)$. The subset $Z(I)$ is identified with the set of maximal ideals containing I. This suggests a way of defining a topology on the set of maximal ideals of any ring. The corresponding idea can also be applied to the set of all prime ideals, and it turns out to be even more useful there. These ideas were first pursued by Oscar Zariski, and the resulting topology bears his name.

Definition. *Let R be any ring. The subsets of Spec R of the form*

$$Z(I) := \{\mathfrak{p} \text{ a prime ideal of } R | \mathfrak{p} \supset I\},$$

for ideals I of R are called ***Zariski-closed subsets****. When there is no danger of confusion, we shall simply call them closed subsets.*

a. Prove that finite unions and arbitrary intersections of closed subsets are closed, and therefore the closed subsets define a topology, called

the **Zariski topology**, on Spec R. The induced topology on the subset max-Spec R is also called the Zariski topology.

If $k = \mathbf{R}$ or $\mathbf{C}$ (or some other topological field), then we have two topologies on k^n: the topology induced from the topology of k, called the **classical topology**, and the Zariski topology. The Zariski topology has many fewer closed sets than the classical topology.

b. Suppose for simplicity that k is an algebraically closed field, in the Zariski topology on $\mathbf{A}^1(k)$ (that is, on the maximal ideals of $k[x]$) show that the open sets are exactly the complements of finite sets. In particular this topology is not Hausdorff. Show that the Zariski topology on $\mathbf{A}^n(k) = k^n$ is *not* the product topology, even for $n = 2$.

c. We define a **distinguished open set** of Spec R to be an open set of the form $U(f) := \{\mathfrak{p} \text{ a prime ideal of } R | f \notin \mathfrak{p}\}$ for some $f \in R$. Show that the distinguished open sets form a basis for the Zariski topology, in the sense that every open set is a union of distinguished open sets. Show that Spec $R = \cup_i U(f_i)$ for some collection f_i of elements of R iff the ideal generated by all the f_i is the unit ideal (1).

d. Show that if R is any ring then Spec R is compact in the Zariski topology (that is, every open covering has a finite refinement).

Exercise 1.25 (max-Spec for rings of continuous functions): (From the formulation of Atiyah and MacDonald [1969], Chapter I, Exercise 26.) Let X be a compact Hausdorff space and let $R = C(X)$ be the ring of continuous real-valued functions on X. Let $\mu : X \to$ max-Spec R be the map taking a point $x \in X$ to the maximal ideal $\mathfrak{m}_x$ of all continuous functions vanishing at x. Prove that μ is a homeomorphism, so that X can be reconstructed algebraically from $C(X)$, as follows:

a. μ is surjective: Let $\mathfrak{m}$ be a maximal ideal of $C(X)$. We wish to prove that $\mathfrak{m} = \mathfrak{m}_x$ for some x. Let $V = V(\mathfrak{m})$ be the set of common zeros of the functions in $\mathfrak{m}$. If V is empty, then for each $x \in X$ there exists $f_x \in \mathfrak{m}$ such that $f_x(x) \neq 0$. Since f_x is continuous, there is an open neighborhood $U(x)$ of x such that f_x does not vanish on $U(x)$. By compactness, X is the union of finitely many of these neighborhoods, say
$$X = U(x_1) \cup \cdots \cup U(x_n).$$
Use these ideas to construct a function $f \in \mathfrak{m}$ that does not vanish anywhere on X. Derive a contradiction.

It now follows that $\mathfrak{m} \subset \mathfrak{m}_x$ for some x. Since $\mathfrak{m}$ is maximal, the ideals are equal.

b. μ is injective: Use Urysohn's lemma (see Kelly [1955], Lemma 4.4; this is the only nontrivial fact required in the proof) to show that if $x \neq y$, then there is a continuous function vanishing at x but not y.

c. μ is a homeomorphism: A subbasis for the topology of X is given by the sets

$$U_f = \{x \in X \mid f(x) \neq 0\}, \quad f \in C(X)$$

while a subbasis for the topology of max-Spec R is given by the sets

$$V_f = \{\mathfrak{m} \in \text{ max-Spec } R \mid f \notin \mathfrak{m}\}, \quad f \in C(X).$$

Show that $\mu(U_f) = V_f$.

2
Localization

A **local ring** is a ring with just one maximal ideal. Ever since Krull's paper [1938], local rings have occupied a central position in commutative algebra. The technique of **localization** reduces many problems in commutative algebra to problems about local rings. This often turns out to be extremely useful: Most of the problems with which commutative algebra has been successful are those that can be reduced to the local case.

Despite this, localization as a general procedure was defined rather late: In the case of integral domains it was described by Grell, a student of Noether's, in [1927], and it was not defined for arbitrary commutative rings until the work of Chevalley [1944] and Uzkov [1948], long after the basic ideas of commutative algebra were established. Perhaps this is because interest was focused on finitely generated algebras on the one hand, and power series rings on the other, and neither of these classes of rings is closed under localization. Instead of passing to a localized ring, as we would now, people often used ideal quotients as a substitute. (We shall explain how this is done in Exercise 2.3.)

The idea of localization, as well as the name, comes from a geometric special case: Given a point p in an algebraic set $X \subset \mathbf{A}_k^r$, we might wish to investigate the nature of X "near" p. That is, we wish to investigate arbitrarily small open neighborhoods of p in the Zariski topology. The Zariski open neighborhoods of p are sets of the form $X - Y$, where Y is an algebraic subset of X not containing p. Now $X - Y$ is generally not isomorphic to an affine algebraic set—for example, the plane minus a point is not (see, for example, Hartshorne [1977], Exercise 3.5). However, small neighborhoods of p in X correspond to large algebraic subsets Y, so we

may assume that Y is the set defined by the vanishing of a single function f, which does not vanish at p. In this case we shall see that $X - Y$ is isomorphic to an algebraic set embedded in $\mathbf{A}_k^{r+1}$, and for this reason we refer to such a set $X - Y$ as an **open affine neighborhood** of p. The affine ring $A(X - Y)$ is obtained from $A(X)$ by adjoining a multiplicative inverse for f; we call this **inverting** f. If we invert all the functions in $A(X)$ not vanishing at p, then the corresponding object, though no longer a finitely generated k-algebra, is a good algebraic representative of the "germ of X at p": It is the **local ring of X at p.**

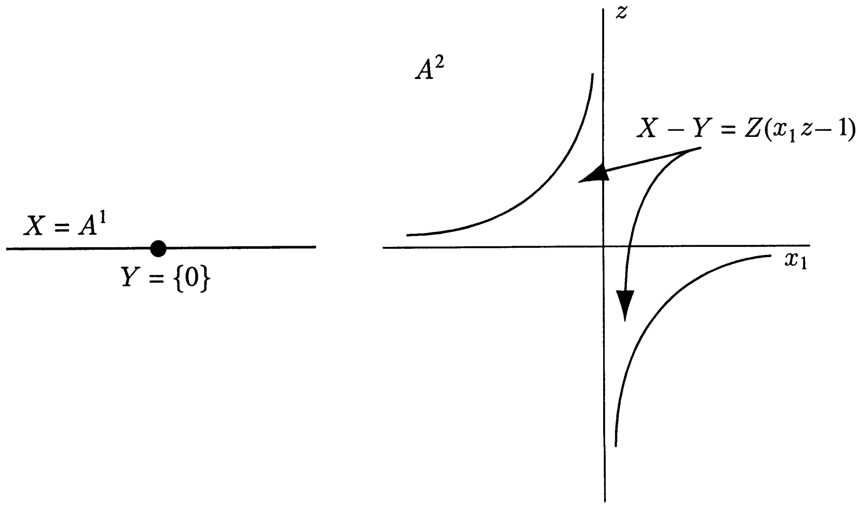
In this chapter we shall explain how to construct new rings from old by inverting arbitrary sets of elements. To motivate the constructions, we return to the problem of removing an algebraic subset Y defined by one equation $f = 0$ from an algebra set X. The points of $X - Y$ are the points x at which $f(x) \neq 0$, so they are the points x such that there is a number, $z(x)$ say, with $z(x)f(x) = 1$. The idea is that $z(x)$ should be a regular function on $X - Y$. If $X \subset \mathbf{A}_k^r$ corresponds to the ideal

$$I \subset k[x_1, \ldots, x_r],$$

then the points of $X - Y$ will correspond—by projection onto the first r coordinates —to the subset of $\mathbf{A}^{r+1}$ defined by the ideal

$$J = I + (zf - 1) \subset k[x_1, \ldots, x_r, z].$$

We may thus define $X - Y$ to be the affine algebraic set in $\mathbf{A}^r \times \mathbf{A}^1 = \mathbf{A}^{r+1}$ corresponding to J, with the inclusion $X - Y \subset X$ given as above by projection to $\mathbf{A}^r$. The following picture gives the simplest case, where we have subtracted $Y = \{0\}$ from $X = \mathbf{A}^1$, and the set $X - Y$ is embedded as a hyperbola in the plane:



In terms of rings, we may write

$$\begin{aligned} A(X - Y) &= k[x_1, \dots, x_r, z]/J \\ &= A(X)[z]/(zf - 1). \end{aligned}$$

Thus we may describe $A(X - Y)$ as the effect of adjoining an inverse of f to $A(X)$ in the "freest" possible way.

2.1 Fractions

In general, we want to be able to put in inverses of many polynomials at once. The product of the inverses of two elements f and g is an inverse for fg, so we shall actually be adjoining the inverses of all the elements of a set U that is **multiplicatively closed**: that is, such that any product of elements of U is in U —including the "empty product" $= 1$. The definition is motivated by the idea of introducing fractions r/u, with $r \in R$ and $u \in U$ (or similarly for elements of a module) as ordered pairs (r, u) modulo the relations that would be satisfied automatically if the elements of U were units and we interpreted r/u as ru^{-1}. There is a mild complication coming from the fact that if $fg = 0$ in R and we adjoin an inverse for f, then we had better make $g = 0$.

Given a ring R, an R-module M, and a multiplicatively closed subset $U \subset R$, we define the **localization of M at U**, written as $M[U^{-1}]$ or $U^{-1}M$, to be the set of equivalence classes of pairs (m, u) with $m \in M$ and $u \in U$ with equivalence relation $(m, u) \sim (m', u')$ if there is an element $v \in U$ such that $v(u'm - um') = 0$ in M. The equivalence class of (m, u) is denoted m/u. We make $M[U^{-1}]$ into an R-module by defining

$$m/u + m'/u' = (u'm + um')/uu' \quad \text{and} \quad r(m/u) = (rm)/u$$

for $m, m' \in M$, $u, u' \in U$, and $r \in R$. Note that $u'm/u'u = m/u$, and the additive inverse of m/u is $(-m)/u$, as one would expect. The localization comes equipped with a natural map of R-modules $M \to M[U^{-1}]$ carrying m to $m/1$.

It is convenient to extend the notation a little further: If $U \subset R$ is an arbitrary set, and $\bar{U} \subset R$ is the multiplicatively closed set of all products of elements in U, then we set $M[U^{-1}] := M[\bar{U}^{-1}]$.

If we apply the definition in the case $M = R$, the resulting localization is a ring, with multiplication defined by

$$(r/u)(r'/u') = rr'/uu',$$

and in fact $M[U^{-1}]$ is an $R[U^{-1}]$-module with action defined by

$$(r/u)(m/u') = rm/uu' \quad \text{for } r \in R, m \in M \text{ and } u, u' \in U.$$

It is useful to have a simple description of when an element localizes to 0:

Proposition 2.1. *Let U be a multiplicatively closed set of R, and let M be an R-module. An element $m \in M$ goes to 0 in $M[U^{-1}]$ (that is, $m/1 = 0$) iff m is annihilated by an element $u \in U$. In particular, if M is finitely generated, then $M[U^{-1}] = 0$ iff M is annihilated by an element of U.*

Proof. The first statement is immediate from the definition. For the second, note that if generators $m_i \in M$ are annihilated by elements $u_i \in U$, then M is annihilated by the product of the u_i. □

As a first example, the quotient field of an integral domain R, which we shall denote by $K(R)$, is the localization $R[U^{-1}]$ where $U = R - \{0\}$. Perhaps the most useful analogue for an arbitrary ring R is to take U to be the set of nonzerodivisors of R, and define the **total quotient ring $\boldsymbol{K(R)}$** of R by $K(R) := R[U^{-1}]$. By Proposition 2.1, $K(R)$ is the "biggest" localization of R such that the natural map $R \to R[U^{-1}]$ is an injection.

The very definition of a prime ideal says that an ideal $P \subset R$ is prime iff $R - P$ is a multiplicatively closed set. Localization at such a multiplicative set is used so often that it has its own notation: If P is a prime ideal and $U = R - P$, then we write R_P for $R[U^{-1}]$. Similarly, for any R-module M, we write M_P for $M[U^{-1}]$. We write $\kappa(P)$ for the ring R_P/P_P, the **residue class field of $\boldsymbol{R}$ at $\boldsymbol{P}$**. For example if R is a domain, so that 0 is a prime ideal, then the quotient field of R is $K(R) = R_0 = \kappa(0)$.

The local ring of an affine variety X at a point $x \in X$ mentioned at the beginning of this chapter may now be defined as follows: If R is the affine coordinate ring of X, and $P \subset R$ is the ideal of functions vanishing at x, then the **local ring of $\boldsymbol{X}$ at $\boldsymbol{x}$**, obtained from R by inverting all the functions that do not vanish at x, is the ring R_P.

If $\varphi : M \to N$ is a map of R-modules, then there is a map of $R[U^{-1}]$-modules $\varphi[U^{-1}] : M[U^{-1}] \to N[U^{-1}]$ that takes m/u to $\varphi(m)/u$, called the **localization of $\boldsymbol{\varphi}$**. This makes localization into a functor from the category of R-modules to the category of $R[U^{-1}]$-modules. Later we shall show (Proposition 2.10) that if M is finitely presented, then every homomorphism $M[U^{-1}] \to N[U^{-1}]$ is a localization. With Exercise 2.10 this establishes a very tight connection between modules over a ring and modules over a localization. Many constructions, such as the direct sum of modules, are preserved by localization (this may be proved directly, and it also follows at once from Lemma 2.4, since tensoring preserves direct sums).

If $\varphi : R \to S$ is any homomorphism of rings with the elements of U going to units, then the elements $\varphi(r)\varphi(u)^{-1} \in S$ must satisfy the same relations as those imposed on the fractions r/u above. Thus, for any such φ there is a uniquely defined extension to a homomorphism $\varphi' : R[U^{-1}] \to S$. This is called the **universal property** of localization; see Figure 2.1. It makes

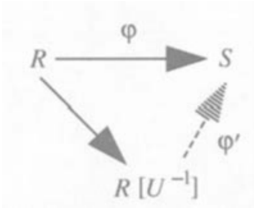


FIGURE 2.1. Universal property: The extension φ' exists (uniquely) iff φ carries elements of U to units.

precise a sense in which $R[U^{-1}]$ is the result of adjoining inverses of elements of U to R in the freest possible way. Another is given in Exercise 2.9.

Notation: For the remainder of this section, R will be a ring, U a multiplicatively closed subset, and M an R-module.

The ideal theory of $R[U^{-1}]$ is a simplified version of the ideal theory of R:

Proposition 2.2. *Let $\varphi : R \to R[U^{-1}]$ be the natural map $r \mapsto r/1$.*

a. *For any ideal $I \subset R[U^{-1}]$ we have $I = \varphi^{-1}(I)R[U^{-1}]$. Thus the map $I \mapsto \varphi^{-1}(I)$ is an injection of the set of ideals of $R[U^{-1}]$ into the set of ideals of R. It preserves inclusions and intersections, and takes prime ideals to prime ideals.*

b. *An ideal $J \subset R$ is of the form $\varphi^{-1}(I)$ for some ideal $I \subset R[U^{-1}]$ iff $J = \varphi^{-1}(JR[U^{-1}])$. This is the case iff each element $u \in U$ is a nonzerodivisor mod J in the sense that if $r \in R$ and $ru \in J$, then $r \in J$. In particular, the correspondence $I \mapsto \varphi^{-1}(I)$ is a bijection between the primes of $R[U^{-1}]$ and the primes of R not meeting U.*

A similar result holds for submodules of an arbitrary module. We leave this easy generalization to the interested reader.

Proof.

a. The inclusion $I \supset \varphi^{-1}(I)R[U^{-1}]$ is obvious and the reverse inclusion follows because for any element $r/u \in I$, with $r \in R$ and $u \in U$, the element r is in $\varphi^{-1}(I)$. It follows at once that $I \mapsto \varphi^{-1}(I)$ is an injection.

   If $\varphi : R \to S$ is any map of sets, then the operation taking subsets of S to subsets of R by $I \mapsto \varphi^{-1}(I)$ preserves inclusions and intersections. If φ is a map of rings and $I \subset S$ is an ideal, then $\varphi^{-1}(I)$ is an ideal of R. Moreover, φ induces an injection $R/\varphi^{-1}(I) \subset S/I$. If, in addition, I is prime then S is a domain, and it follows that $R/\varphi^{-1}(I)$ is a domain, so $\varphi^{-1}(I)$ is prime.

b. If $J = \varphi^{-1}(I)$ then $JR[U^{-1}] \subset I$, so $J = \varphi^{-1}(JR[U^{-1}])$. Since the elements of U act as units on $R[U^{-1}]/I$, they act as nonzerodivisors on

the R-submodule R/J, so J satisfies the given condition. Conversely, suppose that the elements of U act as nonzerodivisors on R/J. If $r \in \varphi^{-1}(JR[U^{-1}])$, then $r/1 \in JR[U^{-1}]$, so $r/1 = j/u$ for some $j \in J$ and $u \in U$. It follows that $uu'r = u'j \in J$ for some $u' \in U$. Since u and u' are nonzerodivisors mod J, we have $r \in J$. Thus $J = \varphi^{-1}(JR[U^{-1}])$, and we are done. The last statement follows because any element not in a prime ideal is a nonzerodivisor modulo that ideal. □

Corollary 2.3. *A localization of a Noetherian ring is Noetherian.*

This is one way in which the Noetherian condition on rings behaves "better" than the condition of being finitely generated over a field.

Proof. If $I \subset R[U^{-1}]$ is an ideal, then by Proposition 2.2, $I = \varphi^{-1}(I)R[U^{-1}]$, so I is generated by the images in $R[U^{-1}]$ of a set of generators of $\varphi^{-1}(I)$. If R is Noetherian, then $\varphi^{-1}(I)$ is finitely generated, so I is too. □

2.2 Hom and Tensor

It is useful and suggestive to express the localization in terms of a more general construction, the **tensor product**. There is a brief treatment of this notion in Appendix A2 (Multilinear Algebra), but we pause here to state some of its main properties. We will often use it along with a closely related construction, the module of homomorphisms between two modules, and we discuss this first.

If M and N are R-modules, then we write $\mathbf{Hom_R(M, N)}$ for the abelian group of all homomorphisms from M to N. It is itself an R-module by the definition

$$(r\varphi)(m) := r\varphi(m) = \varphi(rm) \quad \text{for } r \in R \text{ and } \varphi \in \mathrm{Hom}_R(M, N).$$

The following properties of Hom are very easy to prove from the definition; the reader who is not familiar with these ideas should probably pause and check them as an exercise:

1. $\mathrm{Hom}_R(R, N) \cong N$ by the map $\varphi \mapsto \varphi(1)$.

2. Hom is **functorial** in the sense that if $\alpha : M' \to M$ and $\beta : N \to N'$ are homomorphisms (note their directions!), then there is an induced homomorphism

   $$\mathrm{Hom}_R(M, N) \to \mathrm{Hom}_R(M', N'); \qquad \varphi \mapsto \beta\varphi\alpha.$$

   This homomorphism is often denoted by $\mathrm{Hom}_R(\alpha, \beta)$; or if β is the identity map of N, by $\mathrm{Hom}_R(\alpha, N)$ (and similarly when α is the identity map of M).

3. Hom takes direct sums in the first variable and direct products in the second variable to direct products, in the sense that

$$\mathrm{Hom}_R(\oplus_i M_i, N) = \Pi_i \,\mathrm{Hom}(M_i, N)$$
$$\mathrm{Hom}_R(M, \Pi_j N_j) = \Pi_j \,\mathrm{Hom}(M, N_j);$$

this just says that giving a map from the direct sum $\oplus_i M_i$ to N is the same thing as giving a map from each M_i to N, and similarly for maps from M to $\Pi_j N_j$.

4. If M is an R-module, then the functor $\mathrm{Hom}_R(M, -)$ preserves kernels in the sense that if $A = \ker(\varphi : B \to C)$, then $\mathrm{Hom}_R(M, A) = \ker(\mathrm{Hom}_R(M, \varphi) : \mathrm{Hom}_R(M, B) \to \mathrm{Hom}_R(M, C))$. This is usually expressed by saying that Hom is a **left-exact functor**, which means that if
$$0 \to A \to B \to C$$
is an exact sequence (such a thing is sometimes called a **left-exact sequence**) and M is any module, then the sequence of maps
$$0 \to \mathrm{Hom}_R(M, A) \to \mathrm{Hom}_R(M, B) \to \mathrm{Hom}_R(M, C),$$
obtained because $\mathrm{Hom}_R(M, -)$ is a functor, is exact. (*Interpretation*: Regarding A as a submodule of B, and B/A as a submodule of C, a nonzero map $M \to A$ composes with the monomorphism $A \to B$ to give a nonzero map $M \to B$; and a map $M \to B$ composed with the map to $M/A \subset C$ gives 0 iff the image of M is contained in $A \subset B$.) Similarly, if
$$A \to B \to C \to 0$$
is an exact sequence (sometimes called a **right-exact sequence**), then
$$0 \to \mathrm{Hom}_R(C, N) \to \mathrm{Hom}_R(B, N) \to \mathrm{Hom}_R(A, N)$$
is exact. These two properties are immediate from the definitions.

It is often necessary to work with bilinear maps: If M, N, and P are R-modules, then a **bilinear map** from $M \times N$ to P is defined to be a map of sets (not a map of modules!) $\psi : M \times N \to P$ satisfying the condition of **bilinearity**:

$$\psi((am + a'm') \times (bn + b'n')) = ab\psi(m \times n) + a'b\psi(m' \times n)$$
$$+ab'\psi(m \times n') + a'b'\psi(m' \times n').$$

Bilinear maps may be interpreted in terms of ordinary maps of R-modules by introducing a new module, the **tensor product** $\boldsymbol{M \otimes_R N}$, which may be defined roughly as the module with just enough relations to define a bilinear map $M \times N \to M \otimes_R N$.

More formally, we define $M \otimes_R N$ to be the module with generators $\{m \otimes n | m \in M, n \in N\}$ and relations

$$(am+a'm')\otimes(bn+b'n') = ab(m\otimes n)+a'b(m'\otimes n)+ab'(m\otimes n')+a'b'(m'\otimes n'),$$

mimicking the condition of bilinearity. Note that in particular we have $r(m \otimes n) = (rm) \otimes n = m \otimes (rn)$. When the ring R is clear from context, we sometimes write $M \otimes N$ for $M \otimes_R N$.

It is obvious from the definition that the map $m \times n \mapsto m \otimes n$ is a bilinear map from $M \times N$ to $M \otimes_R N$. Thus, if $\varphi : M \otimes_R N \to P$ is a homomorphism, then the map $\psi : M \times N \to P$ defined by $\psi(m \times n) = \varphi(m \otimes n)$ is bilinear. Conversely, since no relations other than the bilinear relations were imposed on $M \otimes_R N$, if $\psi : M \times N \to P$ is bilinear then there is a unique homomorphism $\varphi : M \otimes_R N \to P$ satisfying $\psi(m \times n) = \varphi(m \otimes n)$.

One point about this construction requires some care: Not every element of $M \otimes_R N$ may be written in the form $m \otimes n$. Rather, every element is expressible (generally in many ways) as a finite sum $\sum m_i \otimes n_i$ for some $m_i \in M$ and $n_i \in N$.

Though brief, the definition of the tensor product is somewhat opaque—for example, it is not easy to tell when two elements $\sum m_i \otimes n_i$ and $\sum m'_j \otimes n'_j$ are equal (though a general criterion is given in Chapter 6). In practice, the following facts are often used to get information about $M \otimes_R N$. The reader will note that they are in a certain sense dual to the preceding facts about Hom. There is a very close relationship between $\otimes$ and Hom, called **adjointness**, explained in Appendix 5. The properties below can be deduced from this relationship, or directly by using the characterization of maps from $M \otimes_R N$ as bilinear maps from $M \times N$, given above.

1. For any module M we have $M \otimes_R R = R \otimes_R M = M$ by isomorphisms sending $1 \otimes m$ and $m \otimes 1$ to m. Also, $M \otimes_R N \cong N \otimes_R M$ by a map sending $m \otimes n$ to $n \otimes m$.

2. The tensor product is **functorial** in the sense that if $\alpha : M' \to M$ and $\beta : N' \to N$ are homomorphisms, then there is an induced homomorphism called $\alpha \otimes \beta : M' \otimes_R N' \to M \otimes_R N$ that sends $m' \otimes n'$ to $\alpha(m') \otimes \beta(n')$.

3. The tensor product preserves direct sums in the sense that if $M = \oplus_i M_i$, then $M \otimes_R N = \oplus_i (M_i \otimes_R N)$.

4. The tensor product preserves cokernels in the sense that if $\alpha : M' \to M$ is a map with cokernel $\operatorname{coker}(\alpha) = M''$, then for any module N the cokernel of the induced map $\alpha \otimes 1 : M' \otimes_R N \to M \otimes_R N$ is $M'' \otimes_R N$. This is usually expressed by saying that the tensor product is **right**

exact in the sense that the functor $-\otimes_R N$ takes an exact sequence of the form

$$M' \to M \to M'' \to 0$$

(that is, a **right-exact sequence**) to an exact sequence

$$M' \otimes_R N \to M \otimes_R N \to M'' \otimes_R N \to 0.$$

These ideas are often used together. For example, to compute the module $M \otimes_R N$ we first find a free presentation of M: That is, we write M as the cokernel of a map of free R-modules $\alpha : \oplus_{i\in I} R \to \oplus_{j\in J} R$, where I and J are sets indexing the bases of the two modules. Giving such a presentation is equivalent to giving generators and relations for M as an R-module. From facts 1 and 3 we see that $(\oplus_{i\in I} R) \otimes_R N = \oplus_{i\in I} N$ is simply a direct sum of copies of N, and similarly for $(\oplus_{j\in J} R) \otimes_R N$. (In fact, if α is written as a matrix using the given bases of the free modules $\oplus_{i\in I} R$ and $\oplus_{j\in J} R$, then the map $\alpha \otimes 1$ is given in a natural sense by the same matrix.) Thus we get an explicit map $\alpha \otimes 1 : \oplus_{i\in I} N \to \oplus_{j\in J} N$ whose cokernel is $M \otimes_R N$. Of course, similar constructions are possible with Hom.

The tensor product is extremely useful in relating the properties of a ring and an algebra over it. If M is an R-module and S is an R-algebra, then $S \otimes_R M$ is not only an R-module, it is also an S-module with multiplication given by the rule $s(t \otimes m) = st \otimes m$ for $s, t \in S$ and $m \in M$. (In the case where S is a localization $R[U^{-1}]$, we shall prove below that $S \otimes_R M = M[U^{-1}]$.)

Carrying this one step further, if A and B are both R-algebras, then the A-module $A \otimes_R B$ is naturally an R-algebra too, with multiplication $(a \otimes b)(c \otimes d) = (ac) \otimes (bd)$. There are natural maps of algebras $A \to A \otimes_R B$ and $B \to A \otimes_R B$ sending $a \mapsto a \otimes 1$ and $b \mapsto 1 \otimes b$. The algebra $A \otimes_R B$ is the "freest" way to put the algebras A and B together into a commutative algebra: Given any commutative R-algebra C and maps $\alpha : A \to C$ and $\beta : B \to C$ of algebras, the map $A \times B \to C$ sending $a \times b$ to $\alpha(a)\beta(b)$ is bilinear, so there is a unique map $A \otimes_R B \to C$ of modules, which turns out to be a map of algebras too, sending $a \otimes b$ to $\alpha(a)\beta(b)$.

The reader who has not seen such things before would do well to pause and try to work out a few examples, such as those given in Exercise 2.4.

There is an amazingly useful geometric interpretation of the tensor product of algebras: If R, A, and B are the coordinate rings of affine varieties X, Y, and Z over an algebraically closed field, then the maps $R \to A$ and $R \to B$ corresponding to the algebra structures correspond to maps $\alpha : Y \to X$ and $\beta : Z \to X$. It turns out that $A \otimes_R B$ is then the coordinate ring of the "fiber product" $\{(y, z) \in Y \times Z | \alpha(y) = \beta(z)\}$. See, for example, Hartshorne [1977, Chapter 2].

The localization of modules can be described in terms of tensor products:

Lemma 2.4. *The natural map* $R[U^{-1}] \otimes_R M \to M[U^{-1}]$ *defined by sending* $r/u \otimes m$ *to* rm/u *is an isomorphism.*

Proof. It is enough to give a map of sets that is inverse to the given map. We define first a map $\alpha : M \times U \to R[U^{-1}] \otimes_R M$ by sending (m, u) to $1/u \otimes m$. We claim that this induces a map $\beta : M[U^{-1}] \to R[U^{-1}] \otimes_R M$. To see that β is defined, suppose that (m', u') is another pair, with $m/u = m'/u'$. This means that there is an element $v \in U$ such that $vu'm = vum'$. Thus $1/(vuu') \otimes vu'm = 1/(vuu') \otimes vum'$. But $1/(vuu') \otimes vu'm = vu'/(vu'u) \otimes m = 1/u \otimes m$ by the definition of the tensor product. Similarly, $1/(vuu') \otimes vum' = 1/u' \otimes m'$. Putting these together gives the desired equality. It is immediate that β is the inverse of the map $r/u \otimes m = 1/u \otimes rm \mapsto rm/u$. □

We next turn to a central property of localization called **flatness**. We say in general that an R-module F is **flat** if for every monomorphism $M' \to M$ of R-modules, the induced map $F \otimes_R M' \to F \otimes_R M$ is again a monomorphism. Since tensor products always preserve right-exact sequences, this is the same as saying that tensoring with F preserves all exact sequences—in particular, it preserves kernels and cokernels. In Chapter 6 we shall explain something of the geometric meaning and uses of this condition. The most interesting case occurs when F is an R-algebra. An example of how this condition is used is given in Proposition 2.10. The next result shows that the condition is satisfied by localizations of R:

Proposition 2.5. *For any multiplicatively closed subset* $U \subset R$*, the ring* $R[U^{-1}]$ *is flat as an* R*-module; that is, localization takes submodules to submodules, and thus preserves kernels and cokernels.*

Proof. Given an injection $M' \subset M$, we must show that

$$R[U^{-1}] \otimes_R M' \to R[U^{-1}] \otimes_R M$$

is an injection. To do this, we use the other description of the localization, and it is enough to prove that the natural map

$$M'[U^{-1}] \to M[U^{-1}]$$

extending the composite

$$M' \to M \to M[U^{-1}]$$

is an injection. But if, for some $m \in M'$, the element m/u goes to zero in $M[U^{-1}]$, then there must be an element $v \in U$ such that $vm = 0$ in M, and this will hold in M' as well. Thus $m/u = 0$ already as an element of $M'[U^{-1}]$, and we are done. □

Here is a useful consequence:

Corollary 2.6. *Localization preserves finite intersections: That is, if* $M_1, \dots, M_t \subset M$ *are submodules, then* $(\cap_j M_j)[U^{-1}] = \cap_j (M_j[U^{-1}])$.

Proof. The point is that intersections can be defined in terms of kernels. Explicitly: The submodule $\cap_i M_i$ is the kernel of the map $M \to \oplus_i M/M_i$. Since localizing preserves kernels, quotients, and direct sums, we see that $(\cap_j M_j)[U^{-1}]$ is the kernel of the map $M[U^{-1}] \to (\oplus_i M/M_i)[U^{-1}] = \oplus_i((M/M_i)[U^{-1}] = \oplus_i M[U^{-1}]/(M_i[U^{-1}])$; that is, $(\cap_j M_j)[U^{-1}] = \cap_j(M_j[U^{-1}])$. □

Unfortunately, localization generally does not preserve infinite intersections (see Exercise 2.5).

We reemphasize the last statement of Proposition 2.1 with a definition: The **support** of M, written **Supp** $\boldsymbol{M}$, is defined to be the set of prime ideals such that $M_P \neq 0$. The last statement of Proposition 2.1 immediately gives:

Corollary 2.7. *If* M *is a finitely generated* R*-module, and* P *is a prime of* R*, then* $P \in \operatorname{Supp} M$ *iff* P *contains the annihilator of* M. □

For those who know about sheaves, the terminology can be explained as follows: In algebraic geometry modules over R are treated as sheaves on Spec R. The stalk of the sheaf corresponding to the module M at the point $P \in$ Spec R is the localization M_P. Support is a well-defined notion for any sheaf; it is the set of points where the stalk of the sheaf is nonzero.

We have already mentioned the geometric interpretation of localization: If X is an affine algebraic set over an algebraically closed field, $R = A(X)$ is its coordinate ring, and $\mathfrak{m} = \mathfrak{m}_p$ is the maximal ideal corresponding to a point $p \in X$, then $R_\mathfrak{m}$ is the ring of "polynomial function germs" on the "germ" of X at p. Here we interpret a germ just as in the theory of manifolds: The germ of a function on the germ of a space X at a point p is by definition the equivalence class of a function defined on some open neighborhood of p, two functions being equivalent if they agree on some (perhaps smaller) open neighborhood on which both are defined. We have already seen that any "polynomial function defined on a neighborhood of p" is of the form f/u for f, $u \in R$ and u not vanishing at p (that is, $u \notin \mathfrak{m}_p$). Two such polynomial functions f/u and g/v represent the same germ if they agree on some small neighborhood of p, which means that they agree on the set wherever some function $w \in R - \mathfrak{m}_p$ is nonzero, that is, $uvw(f/u - g/v) = 0$ as a function on X; this matches exactly the criterion for f/u and g/v to be equal in $R_\mathfrak{m}$. (In the theory of schemes it is convenient to let the definitions go the other way: The germ of X at p is defined in terms of the local ring $R_\mathfrak{m}$; see Hartshorne [1977] or Eisenbud and Harris [1992].)

The statement that a function is zero iff it is zero locally at any point has as its analogue the following extremely useful lemma.

Lemma 2.8. *Let R be a ring and let M be an R-module.*

a. *If $m \in M$, then $m = 0$ iff m goes to zero in each localization $M_{\mathfrak{m}}$ of M at a maximal ideal $\mathfrak{m}$ of R. Similarly,*

b. *$M = 0$ iff $M_{\mathfrak{m}} = 0$ for each maximal ideal $\mathfrak{m}$ of R.*

Proof. m goes to zero in a localization $M_{\mathfrak{m}}$ iff the annihilator I of m is not contained in $\mathfrak{m}$. But $m = 0$ iff $I = R$ iff I is not contained in any maximal ideal of R (this last step uses Zorn's lemma, in general, though for Noetherian rings the existence of a maximal ideal containing a given ideal is of course axiomatic). This proves statement a.

To deduce statement b, note that $M = 0$ iff every element of M is 0; using part a, we see that this happens iff every element of M goes to zero in every localization $M_{\mathfrak{m}}$ at a maximal ideal iff every such $M_{\mathfrak{m}}$ is 0. □

This lemma can be used with Proposition 2.5 to reduce many questions to the local case. Here is a typical step in the reduction process:

Corollary 2.9. *If $\varphi : M \to N$ is a map of R-modules, then φ is a monomorphism (or epimorphism, or isomorphism) iff for every maximal ideal $\mathfrak{m}$ of R the localized map*

$$\varphi_{\mathfrak{m}} : M_{\mathfrak{m}} \to N_{\mathfrak{m}}$$

is a monomorphism (or epimorphism, or isomorphism).

Proof. φ is a monomorphism iff $\ker \varphi = 0$. Because localization is flat, $(\ker \varphi)_{\mathfrak{m}} = \ker(\varphi_{\mathfrak{m}})$. Applying the lemma, we get the first version. The statement about epimorphism is similar (but does not require flatness). The statement about isomorphism is made by putting the first two statements together. □

An easy but useful application is the general form of the "Chinese remainder theorem" given in Exercise 2.6.

We next turn to a more sophisticated bond between a ring and its localizations: Homomorphisms between localizations of nice modules all come from homomorphisms between the original modules. In fact this relation depends only on flatness. The hypothesis we need involves a certain R-module M being finitely presented. In the cases of primary interest to us, R will be Noetherian. In this case M is finitely presented iff M is finitely generated, since if $\varphi : G \to M$ is a surjection from a finitely generated free module G, then $\ker \varphi$, as a submodule of G, is also finitely generated, so M has a finite free presentation.

Proposition 2.10. *Let R be a ring and let S be an R-algebra. If M and N are R-modules, then there is a unique S-module homomorphism*

$$\alpha_M : S \otimes_R \mathrm{Hom}_R(M, N) \to \mathrm{Hom}_S(S \otimes_R M, S \otimes_R N)$$

that takes an element $1 \otimes \varphi \in S \otimes_R \mathrm{Hom}_R(M, N)$ to the S-module homomorphism $1 \otimes_R \varphi : S \otimes_R M \to S \otimes_R N$ in $\mathrm{Hom}_S(S \otimes_R M,\ S \otimes_R N)$. If S is flat over R and M is finitely presented, then α_M is an isomorphism. In particular, if M is finitely presented, then $\mathrm{Hom}_R(M, N)$ localizes in the sense that the map α provides a natural isomorphism

$$\mathrm{Hom}_{R[U^{-1}]}(M[U^{-1}], N[U^{-1}]) \cong \mathrm{Hom}_R(M, N)[U^{-1}]$$

for any subset $U \subset R$.

Proof. By Proposition 2.5, the last statement is a special case of the first statements.

The map of sets $\alpha' : \mathrm{Hom}_R(M, N) \to \mathrm{Hom}_S(S \otimes M, S \otimes N)$ taking a homomorphism φ to the homomorphism $1 \otimes \varphi$ is easily seen to be a map of R-modules. Since the target is an S-module, α' extends to a unique map $\alpha = \alpha_M$ of S-modules with the desired property.

Now we suppose that S is flat and M is finitely presented, and prove that α_M is an isomorphism. First suppose that $M = R$. We may identify $\mathrm{Hom}_R(R, N)$ with N by taking a map φ to the element $\varphi(1)$. Also, $S \otimes_R R = S$, and the same remark shows that $\mathrm{Hom}_S(S \otimes_R R,\ S \otimes_R N) = S \otimes_R N$. It is easy to see that the map $\alpha_R : S \otimes_R N \to S \otimes_R N$ is the identity map.

Next suppose that $M = \oplus_1^m R$ is a free module of finite rank, the direct sum of m copies of R. The functors Hom and $\otimes$ both commute with finite direct sums, and the map α_M also decomposes as a direct sum, $\alpha_{\oplus_1^m R} = \oplus_1^m \alpha_R$. Since each α_R is an isomorphism, so is α_M.

Finally, suppose M is any finitely presented module. Choose a finite free presentation

$$F \xrightarrow{\varphi} G \xrightarrow{\psi} M \to 0.$$

If we tensor with S, then because tensoring preserves right-exact sequences, and because $S \otimes_R F$ and $S \otimes_R G$ are finitely generated free S-modules, we get a finite free presentation of $S \otimes_R M$ as an S-module.

To simplify the notation, we denote the tensor product $S \otimes_R M$ by M', and similarly for other modules and maps. Applying $\mathrm{Hom}_R(-, N)$ to the free presentation of M and $\mathrm{Hom}_S(-, N')$ to that of M', we obtain exact sequences $0 \to \mathrm{Hom}_R(M, N) \to \mathrm{Hom}_R(G, N) \to \mathrm{Hom}_R(F, N)$ and $0 \to \mathrm{Hom}_S(M', N') \to \mathrm{Hom}_S(G', N') \to \mathrm{Hom}_S(F', N')$. Because S is flat over R, we may tensor the first of these with S and still have an exact sequence, $0 \to \mathrm{Hom}_R(M, N)' \to \mathrm{Hom}_R(G, N)' \to \mathrm{Hom}_R(F, N)'$. The map defined in the proposition is, with this notation, a map $\alpha_M : \mathrm{Hom}_R(M, N)' \to \mathrm{Hom}_S(M', N')$; we wish to show that α_M is an isomorphism. By the arguments above, α_F and α_G are isomorphisms. Putting

these maps together, we get a commutative diagram with exact rows

$$\begin{array}{ccccccc}
0 \to & \mathrm{Hom}_R(M,N)' & \xrightarrow{\varphi^{\vee\prime}} & \mathrm{Hom}_R(G,N)' & \xrightarrow{\psi^{\vee\prime}} & \mathrm{Hom}_R(F,N)' \\
& \Big\downarrow \alpha_M & & \Big\downarrow \alpha_G & & \Big\downarrow \alpha_F \\
0 \to & \mathrm{Hom}_S(M',N') & \xrightarrow[\varphi'^{\vee}]{} & \mathrm{Hom}_S(G',N') & \xrightarrow[\psi'^{\vee}]{} & \mathrm{Hom}_S(F',N'),
\end{array}$$

where $\varphi^{\vee\prime}$ and $\psi^{\vee\prime}$ are the maps induced by φ and ψ, and $\varphi'^{\vee}$ and $\psi'^{\vee}$ are the maps induced by φ' and ψ'.

It now follows formally that α_M is an isomorphism, for example, from the "Five-Lemma" of Appendix 3, Exercise A3.11 (one adds another $0 \to$ to the left of each of the rows above to make each one a five-term exact sequence). Here is a direct proof:

First we prove that α_M is a monomorphism. Suppose that $x \in \ker \alpha_M$. Since the diagram commutes, we have $\alpha_G \varphi^{\vee\prime}(x) = \varphi'^{\vee}\alpha_M(x) = 0$. Since α_G is an isomorphism, $\varphi^{\vee\prime}(x) = 0$. Since $\varphi^{\vee\prime}$ is a monomorphism, $x = 0$.

Finally, we show that α_M is an epimorphism. Suppose that $y \in \mathrm{Hom}_S(M', N')$. Since α_G is an epimorphism, we may choose $z \in \mathrm{Hom}_R(G, N)'$ so that $\alpha_G(z) = \varphi'^{\vee}(y)$. To show that z comes from an element of $\mathrm{Hom}_R(M, N)'$, it suffices, since the first row of the diagram is exact, to show that $\psi^{\vee\prime}(z) = 0$. By commutativity, $\alpha_F\psi^{\vee\prime}(z) = \psi'^{\vee}\alpha_G(z) = \psi'^{\vee}\varphi'^{\vee}(y) = 0$. Since α_F is an isomorphism, we see that $\psi^{\vee\prime}(z) = 0$, so $z = \varphi^{\vee\prime}(x)$ for some $x \in \mathrm{Hom}_R(M, N)'$. Furthermore, $\varphi'^{\vee}\alpha_M(x) = \alpha_G\varphi^{\vee\prime}(x) = \alpha_G(z) = \varphi'^{\vee}(y)$. Since $\varphi'^{\vee}$ is a monomorphism, we see that $\alpha_M(x) = y$ as required. □

We shall apply this result often, starting in Theorem 2.13 below. See Exercises 3.3, 4.11–4.13, and 19.4 for some surprising applications beyond the ones in the text.

2.3 The Construction of Primes

The complement of a prime ideal is, as we have already mentioned, a multiplicatively closed subset. There is a sort of converse:

Proposition 2.11. *If R is any commutative ring, $U \subset R$ a multiplicatively closed subset, and $I \subset R$ an ideal maximal among those not meeting U, then I is prime.*

Quite generally, ideals maximal with respect to some property have an uncanny tendency to be prime—see the problems for some more examples, one of which is at the center of the theory of primary decomposition, treated in Chapter 3.

Proof. If f, $g \in R$ are not in I, then, by the maximality of I, both $I + (f)$ and $I + (g)$ meet U. Thus, there are elements of the form $af + i$ and $bg + j$ in U with i, $j \in I$. If fg were in I, then the product of $af + i$ and $bg + j$ would be in I, contradicting the fact that I doesn't meet U. □

Here is a variant of the proof just given that makes the relation to localization obvious: Since distinct ideals of $R[U^{-1}]$ contract to distinct ideals of R, the ideal $IR[U^{-1}]$ must be a maximal ideal, and thus prime. If P is the preimage of $IR[U^{-1}]$ in R, then P is prime. But $I \subset P$, and P does not meet U, so $I = P$.

Note that, for any given U, we can use Zorn's lemma to produce an ideal I as in the proposition.

This simple idea is extremely fruitful. For example, it gives a formula for the radical of an ideal (recall from Chapter 1 that if $I \subset R$ is an ideal, then $\text{rad } I = \{f \in R | f^n \in I \text{ for some } n\}$):

Corollary 2.12. *If I is an ideal in a ring R, then* $\text{rad}\, I = \{f | f^n \in I$ *for some* $n\} = \bigcap_{P \text{a prime containing } I} P$. *In particular, the intersection of all primes of R is the radical of* (0), *which is the set of all nilpotent elements of R.*

Proof. The set rad I is obviously contained in the right-hand side. Conversely, if f is not in rad I, then an ideal maximal among those containing I and disjoint from $\{f^n | n \geq 1\}$ is prime, so f is not contained in the right-hand side. □

2.4 Rings and Modules of Finite Length

Recall that a ring is called Artinian if it satisfies a condition dual to the Noetherian condition: the descending chain condition on ideals. That is, R is Artinian if every descending chain of ideals is finite. We shall see below that any Artinian ring is automatically Noetherian.

We shall show in particular that all the prime ideals in a Noetherian ring R are maximal iff R is Artinian, and in this case there are only finitely many maximal ideals. As a consequence, we shall see that an algebraic set whose coordinate ring is Artinian has only finitely many points (the converse is easy). This is the germ of a fundamental finiteness principle in algebraic geometry.

We shall analyze the structure of Artinian rings and modules over them in terms of localization. Consider a simple example, the ring $\mathbf{Z}/(12)$. It is Artinian (as is every finite ring!) and has maximal ideals (2) and (3). From an elementary course in algebra the reader will know that $\mathbf{Z}/(12) \cong \mathbf{Z}/(4) \times \mathbf{Z}/(3)$—this is a case of the "Chinese remainder theorem," known to Sun-Tsu in the first century A.D. We may give a sophisticated description of the

isomorphism as follows: If we localize at the prime (2), then all the integers not divisible by 2 become units. Thus $(12)_{(2)} = (4)_{(2)}$. On the other hand, the odd numbers are already units in $\mathbf{Z}/(4)$, so $(\mathbf{Z}/(12))_{(2)} = (\mathbf{Z}/(4))_{(2)} = \mathbf{Z}/(4)$. The localization map $\mathbf{Z}/(12) \to (\mathbf{Z}/(12))_{(2)} = \mathbf{Z}/(4)$ sending $\bar{n} \in \mathbf{Z}/(12)$ to $\bar{n}/1$ is the same as the projection map sending $\bar{n} = n + (12)$ to $n + (4)$. Similarly, $(\mathbf{Z}/(12))_{(3)} = \mathbf{Z}/(3)$ by the projection map. Putting these maps together, we get the isomorphism $\mathbf{Z}/(12) \to \mathbf{Z}/(4) \times \mathbf{Z}/(3)$. We shall prove that something similar happens for any Artinian ring.

We begin with a general study of modules with finite composition series: If M is a module, then a **chain** of submodules of M is a sequence of submodules with strict inclusions

$$M = M_0 \supset M_1 \supset \cdots \supset M_n.$$

Such a chain is said to have length n (the number of links). The chain is said to be a **composition series** if each M_j/M_{j+1} is a nonzero simple module (that is, has no nonzero proper submodules). Equivalently, a composition series is a maximal chain of submodules of M. We define **length M** to be the least length of a composition series for M, or ∞ if M has no finite composition series. We shall prove that every composition series for M has the same length.

Of course, a simple module must be generated by any nonzero element, so each $M_j/M_{j+1} \cong R/P$ for some ideal P, which may be described by $P = \operatorname{ann} M_j/M_{j+1}$. Again because M_j/M_{j+1} is simple, P must be a maximal ideal.

The next result, which tells something of the structure of modules of finite length, includes the Jordan-Hölder theorem for modules and the Chinese remainder theorem. (A more usual form of the Chinese remainder theorem, proved with the same methods, is given as Exercise 2.6.)

Theorem 2.13. *Let R be a ring, and let M be an R-module. M has a finite composition series iff M is Artinian and Noetherian. If M has a finite composition series $M = M_0 \supset M_1 \supset \cdots \supset M_n = 0$ of length n, then:*

a. *Every chain of submodules of M has length $\leq n$, and can be refined to a composition series.*
b. *The sum of the localization maps $M \to M_P$, for P a prime ideal, gives an isomorphism of R-modules*

$$M \cong \oplus_P M_P,$$

where the sum is taken over all maximal ideals P such that some $M_i/M_{i+1} \cong R/P$. The number of M_i/M_{i+1} isomorphic to R/P is the length of M_P as a module over R_P, and is thus independent of the composition series chosen.

c. *We have $M = M_P$ iff M is annihilated by some power of P.*

Proof. First suppose that M is Artinian and Noetherian, so that it satisfies both ascending chain condition and descending chain condition on submodules. By the ascending chain condition we may choose a maximal proper submodule M_1, a maximal proper submodule M_2 of M_1, and so on. By the descending chain condition this sequence of submodules must terminate, and it can only terminate when some $M_n = 0$. In this case $M = M_0 \supset M_1 \supset \cdots \supset M_n = 0$ is a composition series for M.

a. Suppose that $M' \subset M$ is a proper submodule. We shall show that length $M' <$ length M. The idea is simple: We intersect the terms of the given composition series for M with M' and derive a shorter composition series for M'.

The quotient $(M' \cap M_i)/(M' \cap M_{i+1})$ is isomorphic to $((M' \cap M_i) + M_{i+1})/M_{i+1} \subset M_i/M_{i+1}$. Since M_i/M_{i+1} is simple, we have either $(M' \cap M_i)/(M' \cap M_{i+1}) = 0$ or else $(M' \cap M_i)/(M' \cap M_{i+1})$ is simple, and $(M' \cap M_i) + M_{i+1} = M_i$.

We claim that the latter possibility cannot happen for every i. Assuming on the contrary that it did, we prove by descending induction on i that $M' \supset M_i$ for every i, and we get a contradiction from the statement $M' \supset M_0 = M$. If $i = n$ then clearly $M \supset M_i$. Supposing by induction that $M' \supset M_{i+1}$, we see that $M' \cap M_i = (M' \cap M_i) + M_{i+1} = M_i$, and it follows that $M' \supset M_i$.

From these facts we see that the sequence of submodules $M' \supset M' \cap M_1 \supset \cdots \supset M' \cap M_n = 0$ can be changed, by leaving out the terms $M' \cap M_i$ such that $M' \cap M_i = M' \cap M_i$, to a composition series for M' whose length is $< n$. Since we could do this for any composition series for M, we get length $M' <$ length M as claimed.

Suppose now that $M = N_0 \supset N_1 \supset \cdots \supset N_k$ is a chain of submodules. We shall show by induction on length M that $k \leq$ length M. This is obvious if length $M = 0$, since then $M = 0$. By the argument above, length $N_1 <$ length M; so by induction, the length of the chain $N_1 \supset \cdots \supset N_k$, is $k - 1$ $\leq$ length N_1. Since length $N_1 <$ length M, it follows that $k \leq$ length M.

From the definition of length it now follows that every maximal chain of submodules has length n, and every chain of submodules can be refined to a maximal chain. Further, n is a uniform bound on the lengths of all ascending or descending chains of submodules, so that M has both ascending chain condition and descending chain condition.

b. By Corollary 2.9 it suffices to show that the given map becomes an isomorphism after localizing at any maximal ideal Q of R. This will be easy once we understand what happens when we localize a module of finite length.

We begin with the case when M has length 1, that is, when M is a simple module. In this case $M \cong R/P$ for some maximal ideal $P = \operatorname{ann} M$. If $P = Q$, then since R/Q is a field, the elements outside of Q act as units on R/Q, and we see that $(R/Q)_Q = R/Q$. If on the other hand $P \neq Q$, then since P is maximal, $P \not\subset Q$, so $P_Q = R_Q$. Thus $(R/P)_Q = R_Q/P_Q = 0$.

It follows in particular from this that if Q and Q' are distinct prime ideals, then $(M_Q)_{Q'} = 0$.

We now return to the general case, length $M = n < \infty$. The composition series for M localizes to a sequence of submodules

$$M_Q = (M_0)_Q \supset (M_1)_Q \supset \cdots \supset (M_n)_Q = 0.$$

The modules M_i/M_{i+1} have length 1, so the case already treated shows that $(M_i/M_{i+1})_Q = M_i/M_{i+1}$ if $Q = \operatorname{ann} M_i/M_{i+1}$, and $(M_i/M_{i+1})_Q = 0$ otherwise. Thus M_Q has a finite composition series corresponding to the subseries of the one for M, obtained by keeping only those $(M_i)_Q$ such that $M_i/M_{i+1} \cong R/Q$. In particular, if none of the modules M_i/M_{i+1} is isomorphic to R/Q, then $M_Q = 0$; and if Q and Q' are distinct maximal ideals, then $(M_Q)_{Q'} = 0$.

Now consider the map $\alpha : M \to \oplus M_P$, the sum of the localization maps, where P ranges over those maximal ideals such that some $M_i/M_{i+1} \cong R/P$. We see from the above that we could harmlessly extend the sum to all maximal ideals; the new terms are all 0. For any maximal ideal Q and any module M we have $(M_Q)_Q = M_Q$, so the identity map is one part of the localization of α:

$$\alpha_Q : M_Q \to (\oplus_{P\text{a maximal ideal}} M_P)_Q = \oplus_{P\text{a maximal ideal}} ((M_P)_Q).$$

But if $P \neq Q$ and M has finite length, then we have seen that $(M_P)_Q = 0$. Thus α_Q is the identity map for every maximal ideal Q, and it follows that α is an isomorphism.

c. Suppose that M is annihilated by a power of a maximal ideal P. If $Q \neq P$ is another maximal ideal, then P contains an element not in Q. This element acts as a unit on M_Q. Since a power of the element acts as 0 on M, we must have $M_Q = 0$. Thus by part b, $M \cong M_P$. Conversely, suppose that $M \cong M_P$. The preceding description of localization shows that every factor $M_i/M_{i+1} \cong R/P$. By induction, we see that $P^d M \subset M_d$, and in particular $P^n M = 0$. □

We now return to Artinian rings. The result that an Artinian ring is Noetherian, which is part of the next theorem, is true even for noncommutative rings (with unit); in the more general setting it is due to Hopkins [1939]. The proof in the commutative case is somewhat simpler. We follow the presentation of Altman and Kleiman [1970].

Theorem 2.14. *Let R be a ring. The following conditions are equivalent:*

a. R is Noetherian and all the prime ideals in R are maximal.

b. R is of finite length as an R-module.

c. R is Artinian.

If these conditions are satisfied, then R has only finitely many maximal ideals.

Proof. a $\Rightarrow$ b: If R is Noetherian and not of finite length, let $I \subset R$ be an ideal maximal with respect to the property that R/I is not of finite length. We claim that I is prime. Indeed, if $ab \in I$ and $a \notin I$, then we may form an exact sequence

$$0 \to R/(I : a) \xrightarrow{a} R/I \to R/(I + (a)) \to 0.$$

Since $I + (a)$ properly contains I, the module $R/(I + (a))$ has finite length. If $b \notin I$, then $(I : a)$ properly contains I as well, so by assumption $R/(I : a)$ also has finite length. Putting together composition series for $R/(I + (a))$ and $R/(I : a)$ we get a composition series for R/I, so it has finite length too, contrary to our assumption. The contradiction shows that $b \in I$. Thus I is prime.

Now suppose in addition that all the prime ideals in R are maximal. If R were not of finite length, then the prime I just constructed would be a maximal ideal and R/I would be a field, contradicting the defining property of I and showing that R is of finite length after all.

b $\Rightarrow$ c: This is clear from Theorem 2.13.

c $\Rightarrow$ a: Suppose that R is Artinian. Our first goal is to show that 0 is a product of maximal ideals of R. Since R is Artinian, we may choose from among all ideals that are products of maximal ideals of R, a minimal such ideal, J. We wish to show that $J = 0$.

For every maximal ideal M of R, the minimality of J implies that $MJ = J$; in particular, $J \subset M$. Since J^2 is also a product of maximal ideals, we have $J^2 = J$. If $J \neq 0$, we can choose an ideal I minimal among ideals not annihilating J. Since $(IJ)J = IJ^2 = IJ \neq 0$, and $IJ \subset I$, we must have $IJ = I$.

Some element $f \in I$ must satisfy $fJ \neq 0$, and since I is minimal, we must have $I = (f)$. Since $IJ = I$, there is an element $g \in J$ such that $f = fg$, or equivalently $(1 - g)f = 0$. Since g is in every maximal ideal, $1 - g$ is in none; that is, $1 - g$ is a unit. Thus $f = 0$. This contradiction shows that indeed $J = 0$.

We now have $0 = M_1 M_2 \cdots M_t$ for some maximal ideals M_i of R. For each s, the quotient $M_1 M_2 \cdots M_s / M_1 M_2 \cdots M_{s+1}$ is a vector space over R/M_{s+1}. Any subspace is a submodule, corresponding to a certain ideal of R containing $M_1 M_2 \cdots M_{s+1}$. Similarly, any descending chain of subspaces

corresponds to a descending chain of ideals of R, and since R is Artinian, any such chain must be finite. Thus $M_1M_2 \cdots M_s/M_1M_2 \cdots M_{s+1}$ is finite dimensional over R/M_{s+1} and has in particular a finite composition series. Putting these composition series together, we see that R has finite length. By Theorem 2.13, R is Noetherian.

Suppose that P is a prime ideal of R. Since $P \supset 0 = M_1M_2 \cdots M_t$, we see that $P \supset M_i$ for some i. Since M_i is a maximal ideal, $P = M_i$, and P is maximal. In particular, every maximal ideal is one of the M_i, so there are only finitely many. □

Applying this result in the geometric context, we get:

Corollary 2.15. *Let X be an affine algebraic set over a field k. The following are equivalent:*

a. X is finite.

b. $A(X)$ is a finite dimensional vector space over k, whose dimension is the number of points in X.

c. $A(X)$ is Artinian.

Proof. a $\Rightarrow$ b: If X is finite, then since $A(X)$ is the ring of polynomial functions restricted to X, we have $A(X) = \Pi_{x \in X} A(x) = \Pi_{x \in X} k$, a direct product of as many copies of the residue field as there are points in X.

b $\Rightarrow$ c: If R is a k-algebra that is finite dimensional as a k-vector space, then any descending chain of subvector spaces is finite, and thus any descending chain of ideals is necessarily finite.

c $\Rightarrow$ a: If $A(X)$ is Artinian, then by Theorem 2.14 it has only finitely many maximal ideals. Since the points of X correspond to maximal ideals, we are done. □

Combining Theorem 2.14 with Theorem 2.13b, we deduce a sort of structure theorem for Artinian rings:

Corollary 2.16. *Any Artinian ring is a finite direct product of local Artinian rings.*

Proof. Since R has finite length as a module over itself, we see from Theorem 2.13 that the sum of the finitely many localization maps, $R \to \oplus_i R_{M_i}$ is an isomorphism of R-modules. The R-algebra $\prod_i R_{M_i}$, which is the direct product of the localizations, is nothing but $\oplus_i R_{M_i}$ when regarded as an R-module. Since each map $R \to R_{M_i}$ is a map of rings, the isomorphism of R-modules $R \to \prod_i R_{M_i}$ is actually an isomorphism of rings, and we see that R is the direct product of finitely many local Artinian rings. □

We can also characterize modules of finite length over Noetherian rings.

Corollary 2.17. *Let R be a Noetherian ring, and let M be finitely generated R-module. The following are equivalent:*

a. M has finite length.

b. Some finite product of maximal ideals $\Pi_{i=1}^n P_i$ annihilates M.

c. All the primes that contain the annihilator of M are maximal.

d. $R/\operatorname{ann}(M)$ is an Artinian ring.

Proof. a $\Rightarrow$ b: If M has finite length, then by Theorem 2.13b and c, M is a direct sum of modules, each of which is annihilated by a power of a certain prime. The product of these powers annihilates M.

b $\Rightarrow$ c: If a product of maximal ideals $\Pi_{i=1}^n P_i$ annihilates M and a prime P contains the annihilator of M, then $P \supset \Pi_{i=1}^n P_i$ and thus $P = P_i$ for some i.

c $\Rightarrow$ d: Immediate from Theorem 2.14.

d $\Rightarrow$ a: Set $S = R/\operatorname{ann}(M)$, and suppose that S is Artinian. By Theorem 2.14, S has finite length as an S-module (or equivalently as an R-module). Since M is a finitely generated S-module, it is a homomorphic image of a finite direct sum of copies of S, and is thus a module of finite length. □

Using Corollary 2.17, we see that every finitely generated module can be made into a module of finite length by localization at a prime minimal over its annihilator.

Corollary 2.18. *Let R be a Noetherian ring, $0 \neq M$ a finitely generated R-module, I the annihilator of M, and P a prime ideal containing I. The R_P-module M_P is a nonzero module of finite length iff P is minimal among primes containing I.*

Proof. If P is a prime ideal minimal among primes containing I, then P_P is nilpotent in R_P/I_P by Corollary 2.12. Thus, a power of P_P annihilates M_P, and Corollary 2.17 shows that M_P has finite length.

Conversely, suppose that M_P has finite length over R_P. The annihilator of M_P is I_P. Thus, by Corollary 2.17, every prime of R_P/I_P is maximal. Since the primes of R_P/I_P correspond to the primes of R containing I and contained in P, we see that P is minimal in the desired sense. □

The most useful special case of these results is where $M = R/I$ (so that in particular $I = \operatorname{ann} M$).

Corollary 2.19. *Let I be an ideal in a Noetherian ring R. The following are equivalent for a prime P containing I:*

a. P is minimal among primes containing I.

b. R_P/I_P is Artinian.

c. In the localization R_P we have $P_P^n \subset I_P$ for all $n \gg 0$.

Proof.

a $\Rightarrow$ b: If P is minimal among primes containing I, then P_P is the unique prime of R_P/I_P. Corollary 2.17 shows that R_P/I_P is Artinian.

b $\Rightarrow$ c: Suppose that R_P/I_P is Artinian. By Theorem 2.14 R_P/I_P has finite length, and by Theorem 2.13c it is annihilated by a power of P_P—that is, $P_P^n \subset I_P$ for large n.

c $\Rightarrow$ a: Suppose that $P_P^n \subset I_P$. If Q is a prime of R such that $I \subset Q \subset P$, then after localizing we see that $P_P^n \subset Q_P$, so $P_P = Q_P$. It follows that $P = Q$. Thus P is minimal among primes containing I. □

2.5 Products of Domains

In a different direction we may use localization to characterize the Noetherian rings that are direct products of domains.

Proposition 2.20. *If R is a Noetherian ring, then R is a finite direct product of domains iff for every maximal ideal P of R, the local ring R_P is a domain.*

Proof. Suppose $R = \prod_i R_i$ is a direct product of domains R_i. A prime ideal P of R cannot contain the unit element e_i of each of the R_i. But if $e_i \notin P$ then since e_i annihilates R_j for $j \neq i$ we have $R_P = (R_i)_P$, a domain.

Conversely, suppose that every localization of R at a maximal ideal is a domain. Let $\{Q_i\}$ be the set of minimal primes of R. Since an intersection of primes in a descending sequence is again prime, this set is nonempty. By the result of exercise 1.2 (or see Theorem 3.1a) there are only finitely many Q_i. We must show that the map $\varphi : R \to \Pi_i R/Q_i$ is an isomorphism. By Corollary 2.9 it is enough to show that φ becomes an isomorphism after localizing at a maximal ideal P of R. The minimal primes of R_P are the localizations of the minimal primes of R contained in P. Since R_P is a domain by hypothesis, there is only one of these, say Q_1, and we have $(Q_1)_P = 0, (Q_i)_P = R$ for $i \neq 1$. It follows that $(\Pi_i R/Q_i)_P = (R/Q_1)_P = R_P$ and φ localizes to an isomorphism as required. □

In trying to prove that a given ring is a domain, using methods of local algebra, one often proves in fact that the ring is locally a domain in the

sense above. Using Proposition 2.20 it is then enough to eliminate the possibility that the ring contains idempotent elements other than 0 and 1. We shall return to these ideas much later, in Theorem 18.15.

2.6 Exercises

Exercise 2.1: Check that the definitions really do make $R[U^{-1}]$ into a ring and $M[U^{-1}]$ into an $R[U^{-1}]$-module (and thus also an R-module). Check that the map $R \to R[U^{-1}]$ sending r to $r/1$ is a ring homomorphism, and the map $M \to M[U^{-1}]$ sending m to $m/1$ is a homomorphism of R-modules.

Exercise 2.2 (An alternate construction of localization): Let R be a ring and let U be any subset of R. Show that $R[U^{-1}]$ is the result of adjoining inverses of elements of U to R in the freest possible way, in the sense that

$$R[U^{-1}] \cong R[\{x_u\}_{u\in U}]/(\{ux_u - 1\}_{u\in U}).$$

Exercise 2.3 (How to localize without admitting it): Here is a collection of results that allow one to do many things in a localization without having to admit that there is any such thing. Suppose $U \subset R$ is a multiplicatively closed subset of a ring. Show that there is a one-to-one correspondence, preserving sums and intersections, between the ideals in $R[U^{-1}]$ and the ideals I in R such that $(I : f) = I$ for all $f \in U$ (recall from Chapter 0 that $(I : f) := \{r \in R | fr \in I\}$). Show that this correspondence respects the property of being prime. Show that for any ideal $J \subset R$ we have $R \cap JR[U^{-1}] = \sum_{f\in U}(J : f^\infty)$ where $(J : f^\infty) := \cup_{n=1}^{\infty}(J : f^n)$. Show that the ideals $I \subset R$ such that $(I : f) = I$ are exactly the image of the map $J \mapsto R \cap JR[U^{-1}]$. Historically, constructions like $(I : f)$ were used before localizations were defined, to accomplish the same ends.

Exercise 2.4 (Practice with Hom and $\otimes$): Let k be a field, and let $\mathbf{Z}$ denote (as usual) the ring of integers. Let m, n be integers. Describe as explicitly as possible:

a. $\mathrm{Hom}_{\mathbf{Z}}(\mathbf{Z}/(n), \mathbf{Z}/(m))$ and $\mathrm{Hom}_{k[x]}(k[x]/(x^n), k[x]/(x^m))$.

b. $\mathbf{Z}/(n) \otimes_{\mathbf{Z}} \mathbf{Z}/(m)$ and $k[x]/(x^n) \otimes_{k[x]} k[x]/(x^m)$.

c. $k[x] \otimes_k k[x]$ (describe this as an algebra).

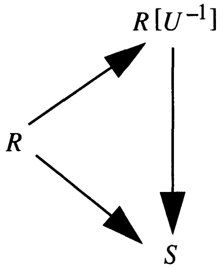
Exercise 2.5: Suppose k is an infinite field, and let U be the set of nonzero elements of the polynomial ring $k[x]$ in one variable. Show that $(\cap_{a\in k}(x - a))[U^{-1}] \neq \cap_{a\in k}((x - a)[U^{-1}])$. Thus, Corollary 2.6 would be false for infinite intersections.

Exercise 2.6 (General form of the Chinese remainder theorem): Let R be a ring, and let $Q_1, \dots, Q_n$ be ideals of R such that $Q_i + Q_j = R$ for all $i \neq j$. Show that $R/(\cap_i Q_i) \cong \Pi_i R/Q_i$ as follows:

a. Consider the map of rings $\varphi : R \to \Pi_i R/Q_i$ obtained from the n projection maps $R \to R/Q_i$. Show that ker $\varphi = \cap_i Q_i$.

b. Let $\mathfrak{m}$ be a maximal ideal of R. Show that the hypothesis that $Q_i + Q_j = R$ for all $i \neq j$ means that at most one of the Q_i is contained in $\mathfrak{m}$. Now use Corollary 2.9 to show that φ is surjective.

Exercise 2.7: Show that the **universal property** of the localization given in the text characterizes $R \to R[U^{-1}]$ up to unique isomorphism in the sense that if another map $R \to S$ has the same property, then there is a unique isomorphism $R[U^{-1}] \to S$ making the diagram commute.



Exercise 2.8: Show that the following universal property similarly characterizes $M \to M[U^{-1}]$: Given a map φ from M to an R-module N on which the elements of U act by multiplication as automorphisms, there is a unique extension $\varphi' : M[U^{-1}] \to N$. In particular, if M and N are $R[U^{-1}]$-modules, then the maps of R-modules from M to N are the same as the maps of $R[U^{-1}]$-modules.

Exercise 2.9: One way of describing the ring $R[U^{-1}]$ is to say what its modules are: Show that an $R[U^{-1}]$-module is the same thing as an R-module on which the elements of U act as automorphisms. In particular, the map $M \to M[U^{-1}]$ is an isomorphism iff the elements of U act as automorphisms on M.

Exercise 2.10: Show that every finitely generated module over $R[U^{-1}]$ is the localization of a finitely generated module over R. Here is a truly trivial statement that sounds deeper: The same is true without the condition finitely generated.

Exercise 2.11:* Let $N' \subset M[U^{-1}]$ be an $R[U^{-1}]$-submodule, and let $N \subset M$ be the preimage of N'. Show that $N' = N[U^{-1}]$.

Exercise 2.12: Show that Proposition 2.10 is sharp in the following sense: Consider the ring $R = \mathbf{Z}$ of integers. Let U be the set of powers of 2. Consider the statement of the proposition with $M = \oplus_{i=1}^{\infty} R$, $N = R$, $S = R[U^{-1}]$.

a. If $N = R$, show that element $\left(\frac{1}{2^i}\right)_{i=1,\ldots,\infty}$ is not in the image of the map $\alpha : (\Pi_{i=1}^{\infty} R)[U^{-1}] \to \Pi_{i=1}^{\infty}(R[U^{-1}])$ of the proposition.

b. Now suppose $N = \oplus_{i=1}^{\infty} R/(2^i)$, with M and S as before. Show that the map sending the generator of the ith factor of M to the generator of the ith factor of N is nonzero in $\operatorname{Hom}_R(M,N)[U^{-1}]$, but goes to 0 under α.

Exercise 2.13 (Splitting criteria for a short exact sequence): Suppose that

$$(*) \qquad 0 \to A \to B \to C \to 0$$

is a short exact sequence of R-modules.

a. Show that $(*)$ is split iff the map $\operatorname{Hom}_R(C,B) \to \operatorname{Hom}_R(C,C)$ induced by the right-hand map of $(*)$ is an epimorphism.

b. Suppose that $(*)$ is **locally split** in the sense that for each maximal ideal $P \subset R$ the localized sequence $0 \to A_P \to B_P \to C_P \to 0$ is split. If C is finitely presented, show that $(*)$ is split by using part a and Proposition 2.10.

$\mathbf{Z}$-graded Rings and Their Localizations

If we invert an element of a graded ring, even a homogeneous element, we usually do not get a graded ring in the sense of Chapter 1: Negative degrees will occur in the obvious grading. Thus we introduce the notion of a $\mathbf{Z}$-graded ring:

Definition. *A* ***Z-graded ring*** *is a ring R such that*

$$R = \cdots R_{-2} \oplus R_{-1} \oplus R_0 \oplus R_1 \oplus R_2 \oplus \cdots$$

as abelian groups and $R_iR_j \subset R_{i+j}$. The elements of R_i are called ***homogeneous*** *elements of degree i. A* ***homogeneous ideal*** *in a $\mathbf{Z}$-graded ring is simply an ideal generated by homogeneous elements.*

The case of ordinary graded rings is the case where $R_i = 0$ for $i < 0$.

Exercise 2.14 (Characterization of homogeneous ideals): Show that an ideal I of a $\mathbf{Z}$-graded ring R is homogeneous iff for every element $f \in I$, all the homogeneous components of f are in I.

Exercise 2.15: Many basic operations on ideals, when applied to homogeneous ideals in **Z**-graded rings, lead to homogeneous ideals. For example, let I be a homogeneous ideal in a **Z**-graded ring R. Show that:

a. The radical of I is homogeneous; that is, the radical of I is generated by all the homogeneous elements f such that $f^n \in I$ for some n.

b. If I and J are homogeneous ideals of R, then

$$(I : J) := \{f \in R | fJ \subset I\}$$

is a homogeneous ideal.

c. Suppose that for all f, g homogeneous elements of R such that $fg \in I$, one of f and g is in I. Show that I is prime.

See Section 3.5 for further results in this direction.

Exercise 2.16: Let R be a **Z**-graded ring and let M be a graded R-module. Show that if x is a homogeneous element of nonzero degree, then $u := 1 - x$ is a nonzerodivisor on M. The element u is a unit iff x is nilpotent.

Given a projective variety $X \subset \mathbf{P}_k^r$, it is very useful to be able to write the localizations of the affine coordinate rings of the affine open pieces of X directly in terms of the homogeneous coordinate ring of X. The following exercise explains how to do this, in a form that works for arbitrary **Z**-graded rings.

Exercise 2.17 (Localization of graded rings): Suppose R is a **Z**-graded ring and $0 \neq f \in R_1$. Show that $R[f^{-1}]$ is again a **Z**-graded ring. Let $S = R[f^{-1}]_0$.

a. Show that $R[f^{-1}] \cong S[x, x^{-1}]$, where x is a new variable. (The ring $S[x, x^{-1}]$ is called the ring of **Laurent polynomials** over S. We make the convention that if S is the zero ring, then $S[x, x^{-1}]$ is also the zero ring.)

b. Show that $S = R[f^{-1}]_0 \cong R/(f - 1)$.

c. Let $U \subset R$ be a multiplicatively closed set of homogeneous elements containing at least one nonzero element of R_1. Show that $R[U^{-1}] \cong (R[U^{-1}])_0[x, x^{-1}]$, where x is an indeterminate of degree 1.

d. Now let P be a homogeneous prime ideal of R, and let U be the multiplicative set of homogeneous elements not in P. Note that $R[U^{-1}]$ is naturally a **Z**-graded ring. We define $R_{(P)}$ to be the degree-0 component $R[U^{-1}]_0$ of $R[U^{-1}]$. If P does not contain R_1, then by part c, $R[U^{-1}] \cong R_{(P)}[x, x^{-1}]$.

Suppose $f \in R_1$, but $f \notin P$. Write Q for the ideal that is the image of P in $R/(f-1)$. Show that Q is a prime ideal and that

$$R_{(P)} = (R/(f-1))_Q.$$

If R is the homogeneous coordinate ring of a projective variety X, and P is the ideal of a subvariety Y, then these objects have a geometric meaning: If V is the affine open subset $x = 1$, for some linear form x, and V meets Y nontrivially, then $R_{(P)}$ is the ordinary localization of the affine coordinate ring of V at the prime ideal $Q = I(V \cap Y)$.

Exercise 2.18: Show that if R is a graded ring with no nonzero homogeneous prime ideals, then R_0 is a field and either $R = R_0$ or $R = R_0[x, x^{-1}]$.

Partitions of Unity

Exercise 2.19 (Partition of unity): Let R be a ring and let M be an R-module. Suppose that $\{f_i\}$ is a set of elements of R that generate the unit ideal. Prove:

a.* If $m \in M$ goes to 0 in each $M[f_i^{-1}]$, then $m = 0$.

b.* If $m_i \in M[f_i^{-1}]$ are elements such that m_i and m_j go to the same element of $M[f_i^{-1}f_j^{-1}]$, then there is an element $m \in M$ such that m goes to m_i in $M[f_i^{-1}]$ for each i. Note that by part a, the element m is unique.

This result is the essential point in establishing that R-modules are sheaves on Spec R. It plays the role of the classical "partition of unity argument" in geometry. For example, in the case $M = R$ it allows one to piece together global functions from functions defined on each open set of a covering and agreeing on the overlaps. See, for example, Eisenbud and Harris [1992].

Exercise 2.20: There are other collections of localizations that have the property of the set of localizations at all maximal ideals described in Corollary 2.9. Perhaps the most important type is the following, which generalizes the covering of an affine set by open affine subsets: Show that for a collection of elements $f_1, \ldots, f_m \in R$, the following properties are equivalent:

a. The ideal generated by $f_1, \ldots, f_m$ is R.

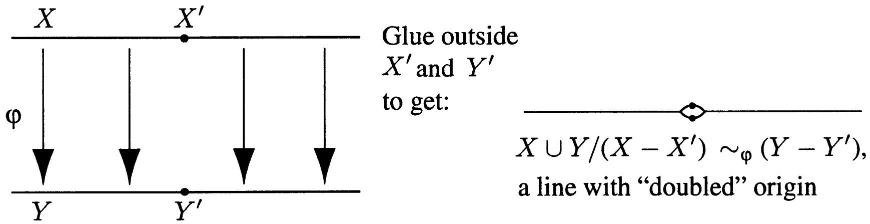
b. An R-module M is zero iff each of the modules $M[f_i^{-1}]$ is zero.

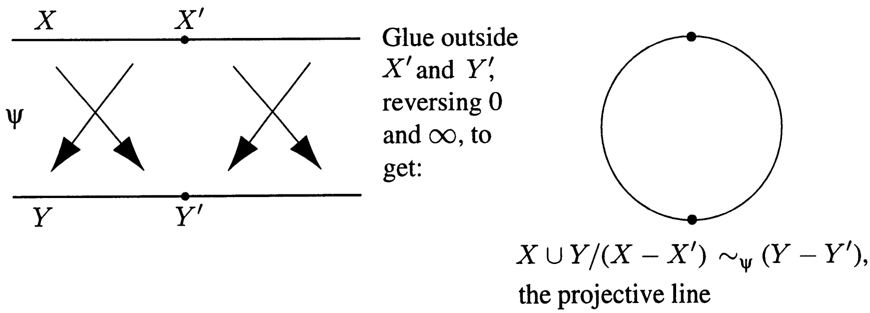
Gluing

An important use of localization in geometry is to construct new algebraic sets by gluing together old ones along open subsets, just as in the classical theory of manifolds. For example, take two copies X and Y of the affine line $\mathbf{A}^1_k$ over k, corresponding to the affine algebras $k[s]$ and $k[t]$. Let $X' \subset X$ and $Y' \subset Y$ be the origins, so the open affine subsets $X - X'$ and $Y - Y'$ correspond to the algebras $k[s, s^{-1}]$ and $k[t, t^{-1}]$, respectively. Clearly, $X - X' \cong Y - Y'$ in (at least) two different ways. We write

$$\begin{aligned} \varphi &: X - X' \to Y - Y' \quad x \mapsto x \\ \psi &: X - X' \to Y - Y' \quad x \mapsto x^{-1} \end{aligned}$$

for two isomorphisms. If we glue together $X - X'$ and $Y - Y'$ by φ, we get a strange, rather nongeometric space, a line with the origin doubled. But if we use ψ as the gluing map, we get the projective line, as the following picture suggests and Exercise 2.21 proves.





Exercise 2.21:

a. Show that $\mathbf{P}^1$ is obtained by gluing two copies of $\mathbf{A}^1$ as follows: Consider the two open subsets $U_0 = \{(a_0, a_1) \in \mathbf{P}^1 | a_0 \neq 0\}$ and $U_1 = \{(a_0, a_1) \in \mathbf{P}^1 | a_1 \neq 0\}$, as described in Chapter 1. Each of these is identified with an affine space: With notation as above we may write the identifications as $U_0 \cong X$ by $(a_0, a_1) \mapsto a_1/a_0$ and $U_1 \cong Y$

by $(a_0, a_1) \mapsto a_0/a_1$. Show that $U_0 \cap U_1$ is taken by these two identifications to $X - X'$ and $Y - Y'$, respectively. Show that the composite identification $X - X' \to U_0 \cap U_1 \to Y - Y'$ is the map $s \mapsto t^{-1}$.

b. Formulate a corresponding "gluing" description of $\mathbf{P}^n$.

Constructing Primes

The next exercises exemplify the tendency of ideals maximal with respect to some property to be prime (see also Proposition 2.11). Exercise 2.24 gives an application.

Exercise 2.22:* (Cohen [1950]): An ideal maximal with respect to not being finitely generated is prime; thus a ring whose primes are finitely generated is Noetherian.

Exercise 2.23: (M. Isaacs): An ideal maximal among those that are not principal is prime.

Exercise 2.24: Let R be a Noetherian ring, and let n be a natural number. Show that there are only finitely many primes P of R such that the cardinality of R/P is $\leq n$ as follows:

a. Suppose that R has infinitely many such primes. Let $I \subset R$ be an ideal maximal among those for which R/I has infinitely many such primes. Show that I must be prime. Replacing R by R/I, we may assume from the outset that R is a domain and that every proper homomorphic image of R satisfies the desired statement.

b. Note that R must be infinite (otherwise R has only finitely many ideals!). Let $a_1, \ldots, a_{n+1}$ be distinct elements of R, and let $p = \Pi_{i<j}(a_i - a_j)$ be the product of their differences. Because R is a domain, $p \neq 0$. If $P \subset R$ is a prime ideal, and $p \notin P$, show that the cardinality of R/P is greater than n. Using the hypothesis at the end of step a, show that there are only finitely many primes P containing p for which the cardinality of R/P is n or less than n.

Idempotents, Products, and Connected Components

Exercise 2.25 (Idempotents and connectedness):* If R is a ring, then as in Exercise 1.25 we write Spec R for the topological space whose points are the prime ideals of R and whose closed sets are the sets of prime ideals containing a given ideal of R. Show that Spec R is **disconnected** —that is, Spec R is the disjoint union of two nonempty closed sets, say X_1, X_2, iff R contains a **nontrivial idempotent**—that is, an element $e \neq 0, 1$ such that $e^2 = e$, as follows.

First, if e is a nontrivial idempotent, show that 1-e is also a nontrivial idempotent, and $e(1-e) = 0$. Take X_1 and X_2 to be the sets of primes containing e and 1-e, respectively. Show that Spec R is the disjoint union of X_1 and X_2 and that these sets are nonempty.

For the converse, suppose that Spec R is the disjoint union of nonempty closed sets X_1 and X_2.

a. Since X_j is closed, there is an ideal I_j such that $P \in X_j$ iff $P \supset I_j$. Show that $I_1 + I_2 = R$ and every element of I_1I_2 is nilpotent.

b. Write $1 = a_1 + a_2$ with $a_j \in I_j$. By part a, a_1a_2 is nilpotent, say $(a_1a_2)^n = 0$. By splitting up the right-hand side of the expression $1 = (a_1+a_2)^{2n} = a_1^{2n} + \cdots + a_2^{2n}$ suitably, show that a_1 and a_2 may be replaced with elements e_1 and e_2 such that $e_1 + e_2 = 1$ and $e_1e_2 = 0$.

c. Show that e_1 and e_2 are nontrivial idempotents.

Exercise 2.26 (Idempotents and products): Show that R can be written as a direct product of two or more (nonzero) rings iff R contains a nontrivial idempotent. Show that if e is an idempotent, then $R = Re \times R(1-e)$, and that Re may be realized as a localization, $Re = R[e^{-1}]$.

Exercise 2.27 (Products and their modules):* If R_γ, $\gamma \in \Gamma$, are rings, then the direct product of rings $\Pi_{\gamma\in\Gamma}R_\gamma$ is the **direct product** of the sets R_γ, with componentwise ring operations. If R is a Noetherian ring all of whose primes P_γ are maximal, then by the Chinese remainder theorem, Proposition 2.14, we have $R = \Pi_\gamma R_\gamma$, and the product is finite. In this case $R_\gamma = R_{P_\gamma}$ is a localization of R. We have seen in Proposition 2.17 that the modules over R are made by taking a direct product of modules, one over each R_{P_i}. These phenomena are somewhat more general:

a. Suppose that $R = \Pi_{\gamma\in\Gamma}R_\gamma$ is a finite direct product of rings. Write $e_\gamma \in R$ for the element whose γth component is 1 and whose δth component is 0 for $\delta \neq \gamma$. Show that every R-module M is uniquely expressible as a direct product $M = \Pi_{\gamma\in\Gamma}M_\gamma$, where M_γ is an R_γ-module and the action is componentwise; show in fact that $M_\gamma = e_\gamma M = M[e_\gamma^{-1}]$. Show that any homomorphism

$$\varphi : M = \prod_{\gamma\in\Gamma} M_\gamma \to N = \prod_{\gamma\in\Gamma} N_\gamma$$

of R-modules is a direct product of homomorphisms $\varphi_\gamma : M_\gamma \to N_\gamma$ of R_γ-modules.

b. Show that if $R = \Pi_{\gamma\in\Gamma}R_\gamma$ is an infinite direct product, then not every module is the product of R_γ-modules as above.

3

Associated Primes and Primary Decomposition

As we have suggested in Chapter 1, the earliest impulse toward the development of what is now commutative algebra came from the desire of the number theorists to make use of unique factorization in rings of integers in number fields other than **Q**. When it became clear that unique factorization did not always hold, the search for the strongest available alternative began. The theory of primary decomposition is the direct result of that search. Given an ideal I in a Noetherian ring R, the theory identifies a finite set of "associated" prime ideals of R, and tells how to "decompose" I as an intersection of "primary" ideals that are closely connected with these prime ideals. More generally, the theory produces such a set of associated primes and a decomposition of any submodule of a finitely generated R-module.

Besides the search for an analog of unique prime factorization, there is another reason why primary decomposition is historically important in commutative algebra. Lasker, who first formulated primary decomposition in [1905], was able to do it only for affine rings and convergent power series rings. His proofs used complicated arguments from elimination theory to make an induction on the number of variables. Emmy Noether rewrote the subject in her brilliant paper [1921]. Here she developed the general theory of primary decomposition from the ascending chain condition alone. In this way she enormously simplified the theory and extended its reach. This paper and her subsequent paper on Dedekind domains [1927] showed the importance of the ascending chain condition. It is this work we honor with the name Noetherian.

To help the reader digest the theory that follows, we begin with three examples.

1. Corresponding to the unique prime factorization

$$n = \pm p_1^{d_1} \cdots p_t^{d_t}$$

of an integer in $\mathbf{Z}$ into powers of distinct primes, we may write the ideal (n) as

$$(n) = (p_1^{d_1}) \cap \cdots \cap (p_t^{d_t}).$$

(*Proof*: By induction on t we have $J := (p_2^{d_2} \cdots p_t^{d_t}) = (p_2^{d_2}) \cap \cdots \cap (p_t^{d_t})$, and it suffices to show that if $I = (p_1^{d_1})$ then $IJ = I \cap J$. If $I,\ J \subset R$ are ideals in any commutative ring, then $IJ \subset I \cap J$, but generally the containment may be strict. However, if $I + J = R$, as in our case, we can write $1 = i + j$ with $i \in I$ and $j \in J$. Thus if $f \in I \cap J$, then $f = 1f = if + jf \in IJ + JI = IJ$, so $I \cap J = IJ$. For a generalization, see Exercise A3.17.)

In this case we shall see that the associated primes of (n) are the primes (p_i), and the primary components of (n) are the ideals $(p_i^{d_i})$. This is the sense in which the theory of primary decomposition generalizes the unique factorization of integers.

2. Consider the geometric setting, where $R = k[x_1, \ldots, x_r]$ is a polynomial ring over an algebraically closed field. An algebraic set X in affine r-space over k is called **irreducible** if it cannot be expressed as the union of two properly smaller algebraic sets. If $I \subset k[x_1, \ldots, x_r]$ is the ideal of X, then X is irreducible iff I is prime.

(*Proof*: If X is irreducible and $fg \in I$, then $Z(I, f) \cup Z(I, g) = X$, so f or g must vanish on X and be in I. Conversely, suppose $X = X_1 \cup X_2$. If each X_i is an algebraic set smaller than X, then there is a function f_i vanishing on X_i but not X. Since $f_1 f_2$ vanishes on X, we have $f_1 f_2 \in I$ though neither f_i is in I.)

If X is any algebraic set, then the ideal I of X is a radical ideal so, by Corollary 2.12, I is an intersection of prime ideals. We shall see that I may be written in a unique minimal way as a finite intersection of primes. The primary decomposition of I is this expression. It corresponds to writing X in a unique minimal way as the union of irreducible algebraic sets X_i. We may think of the decomposition as specifying I to be the set of polynomials that vanish on each of the X_i.

3. The ideal $I := (x^2, xy) \subset k[x, y]$ may be written as

$$I = (x) \cap (x^2, xy, y^2),$$

and described as the ideal of polynomials vanishing along the line $x = 0$ *and* vanishing to order at least two at the point $x = y = 0$. Note that the given decomposition is not unique: We could also write $I = (x) \cap (x^2, y)$, which corresponds to saying that a polynomial f is in I if it vanishes along the line $x = 0$ and its derivative $\partial f/\partial x$ vanishes at the point $x = y = 0$.

In this case we shall see that the associated primes of I are the primes (x) and (x, y). The primary component of I corresponding to the prime (x) is (x), while the primary component corresponding to (x, y) is not uniquely defined, but may be taken to be either (x^2, xy, y^2) or (x^2, y).

Quite generally, given any ideal $I \subset k[x_1, \ldots, x_r]$ with k algebraically closed, primary decomposition theory produces a finite set of irreducible algebraic sets X_i—possibly with some embedded in others—and says that I can be specified as the set of polynomial functions satisfying certain "vanishing conditions" on the X_i.

3.1 Associated Primes

Let R be a ring and let M be an R-module.

Definitions. *A prime P of R is* ***associated*** *to M if P is the annihilator of an element of M. The set of all primes associated to M is written* $\mathbf{Ass}_R M$ *or simply* $\mathbf{Ass} M$ *when there can be no confusion.*

Tradition dictates one exception to this terminology: If I is an ideal of R, then the associated primes of the module R/I are called associated primes of I. Confusion rarely arises in this way, since the associated primes of I as a module are usually not interesting. For example, if R is a domain and I is a nonzero ideal, then the only associated prime of the module I is 0.

From the definition we see that P is an associated prime of M iff R/P is isomorphic to a submodule of M. Note that all the associated primes of M contain the annihilator of M.

The next Theorem gathers the central results about associated primes.

Theorem 3.1. *Let R be a Noetherian ring and let M be a finitely generated, nonzero R-module.*

a. $\operatorname{Ass} M$ *is a finite, nonempty set of primes, each containing ann M. The set* $\operatorname{Ass} M$ *includes all the primes minimal among primes containing ann M.*

b. *The union of the associated primes of M consists of* 0 *and the set of zerodivisors on M.*

c. *The formation of the set* Ass M *commutes with localization at an arbitrary multiplicatively closed set* U*, in the sense that*

$$\mathrm{Ass}_{R[u^{-1}]} M[U^{-1}] = \{PR[U^{-1}] | P \in \mathrm{Ass}\, M \text{ and } P \cap U = \emptyset.\}$$

The proof will be given after a series of preliminary results and corollaries. Essentially because of the second part of conclusion a, the primes minimal among those primes containing a given ideal I appear rather often in what follows. To simplify our language, we usually call them **primes minimal over *I***.

The primes in Ass M that are not minimal are called **embedded** primes of M. If $M = R/I$ corresponds to a subscheme $X = \mathrm{Spec}\, R/I$ of Spec R, then the varieties associated to minimal primes over I are called **isolated components** of X, and the varieties associated to other associated primes are called **embedded components** of X (geometrically, they are "embedded in" the isolated components).

If R is a graded Noetherian ring and M is a finitely generated, graded R-module, then the associated primes of R are homogeneous, as we shall see in Proposition 3.12. This allows one to make graded versions of Theorem 3.1 and all the other results in this chapter.

One important consequence of Theorem 3.1 is as follows.

Corollary 3.2. *Let* R *be a Noetherian ring and let* M *be a finitely generated, nonzero* R*-module. Every ideal consisting entirely of zerodivisors on* M *actually annihilates some element of* M*.*

To prove this we need to know that an ideal contained in a union of primes is contained in one of them. This somewhat surprising but elementary fact often goes under the name *prime avoidance.*

3.2 Prime Avoidance

Lemma 3.3 (Prime Avoidance). *Suppose that* $I_1, \ldots, I_n, J$ *are ideals of a ring* R*, and suppose that* $J \subset \cup_j I_j$*. If* R *contains an infinite field or if at most two of the* I_j *are not prime, then* J *is contained in one of the* I_j*.*

If R *is graded,* J *is generated by homogeneous elements of degree* > 0*, and all the* I_j *are prime, then it is enough to assume that the homogeneous elements of* J *are contained in* $\cup_j I_j$*.*

Despite the odd hypotheses, the lemma is rather sharp; see Exercise 3.17. The name "prime avoidance" comes from the following typical application: If an ideal I is not contained in any of a finite number of primes P_j, then there is an element of I that "avoids" being contained in any of the P_j. In the geometric setting we can translate this by saying that if a finite

number of subvarieties X_j of a variety X are given, along with polynomial functions $f_1, \ldots, f_s$ on X, not all vanishing on any of the X_j, then there is some polynomial linear combination $f = \Sigma g_i f_i$ that does not vanish on any of the X_j. The last part will be used in Chapter 14. In fact, the first of the g_i can often be chosen to be 1; see Exercise 3.19 for this and a refinement, and see McAdam [1974] for further refinements and a history of the ring-theoretic formulations of this result.

Proof of Lemma 3.3. If R contains an infinite field, the result is trivial: No vector space over an infinite field can be a finite union of proper subspaces.

In the other case, we do induction on n; the case $n = 1$ is trivial. By induction we may suppose that J is not contained in any smaller union of the I_j, so we can find elements $x_i \in J, x_i$ not in $\cup_{j \neq i} I_j$. Supposing that $J \subset \cup I_j$, we must have $x_i \in I_i$.

If $n = 2$, then $x_1 + x_2$ is in neither I_1 nor I_2, contradicting the supposition. If on the other hand $n > 2$, then we may assume that I_1 is prime, and $x_1 + x_2 x_3 \cdots$ is not in any of the I_j, again a contradiction.

For the graded case we can use the same proof after raising the x_i to a power, chosen so that x_1 and the product $x_2 x_3 \cdots$ have the same degree. We need the hypothesis that each I_j is prime to ensure that for each j the powers of x_i are not in I_j for $j \neq i$. □

Note that in Lemma 3.3 we did not assume that R was Noetherian; we shall have occasion to use the result in a (possibly) non-Noetherian case in Proposition 13.10. Also, in the cases not involving a ground field, the proof just given uses only that J is a subring—without unit—of R.

Proof of Corollary 3.2. By Theorem 3.1 an ideal consisting of zerodivisors on M is contained in the union of the associated primes of M. By Lemma 3.3, it is in one of them. □

Theorem 3.1 clearly implies that if M is nonzero, then $\operatorname{Ass} M$ is nonempty. For example, since the intersection of a descending chain of primes is certainly prime, there are (even without Noetherian hypotheses) always primes minimal over a given ideal. The first step in the proof is to establish the existence of an associated prime directly.

Proposition 3.4. *Let R be a ring and let M be an R-module. If I is an ideal of R maximal among all ideals of R that are annihilators of elements of M, then I is prime (and thus belongs to* $\operatorname{Ass} M$*). In particular, if R is a Noetherian ring, then* $\operatorname{Ass} M$ *is nonempty.*

Proof. If $rs \in I$ and $s \notin I$, then we must show that $r \in I$. If $m \in M$ is an element with $\operatorname{ann} m = I$, then $rsm = 0$ but $sm \neq 0$. Thus (r, I) is contained in the annihilator of sm, and since I was maximal, $(r) + I = I$. Thus $r \in I$. □

Proposition 3.4 is the basis for one of the characteristic applications of the theory of associated primes. If $x \in M$ is an element of any module over any (not necessarily Noetherian) ring R, then by Lemma 2.8 we can test whether $x = 0$ by seeing whether x goes to 0 in the localization M_P for each prime, or even each maximal ideal P. Now we see that if R is Noetherian we can restrict our attention to the associated primes. If M is finitely generated there will be only finitely many of these, a great improvement.

Corollary 3.5. *Suppose that M is a module over a Noetherian ring R.*

a. *If $m \in M$, then $m = 0$ iff m goes to 0 in M_P for each of the maximal associated primes of M.*

b. *If $K \subset M$ is a submodule, then $K = 0$ iff $K_P = 0$ for all $P \in \operatorname{Ass} M$.*

c. *If $\varphi : M \to N$ is a homomorphism from M to an R-module N, then φ is a monomorphism iff the localization $\varphi_P : M_P \to N_P$ is a monomorphism for each associated prime P of M.*

Proof.

a. Suppose $m \neq 0$. Since R is Noetherian, there is a prime maximal among the annihilators of elements of M that contain ann m, and this prime is an associated prime of M by Proposition 3.4. Thus ann m is contained in a maximal associated prime P, so $m/1 \neq 0$ in M_P.

b. If $K = 0$ then clearly $K_P = 0$ for all P. If $K \neq 0$, choose $0 \neq m \in K$ and apply part a.

c. By Proposition 2.5, $(\ker \varphi)_P = \ker(\varphi_P)$. The result follows by putting $K = \ker \varphi$ in part b. □

Proposition 3.4 makes the proof of part b of Theorem 3.1 immediate: If r annihilates a nonzero element of M, then r is contained in a maximal annihilator ideal.

To prove part a we shall apply the following tool:

Lemma 3.6.

a. *If $M = M' \oplus M''$, then $\operatorname{Ass} M = (\operatorname{Ass} M') \cup (\operatorname{Ass} M'')$.*

b. *More generally, if $0 \to M' \to M \to M'' \to 0$ is a short exact sequence of R-modules, then $\operatorname{Ass} M' \subset \operatorname{Ass} M \subset (\operatorname{Ass} M') \cup (\operatorname{Ass} M'')$.*

Proof.

a. Given part b, it is enough to observe that $\operatorname{Ass} M'' \subset \operatorname{Ass} M$.

b. The first containment is clear from the definition. For the second, suppose that $P \in \operatorname{Ass} M - \operatorname{Ass} M'$. If $x \in M$ has annihilator P, so

that $Rx \cong R/P$, then since P is prime every nonzero submodule of Rx also has annihilator P. It follows that $Rx \cap M' = 0$, so Rx is isomorphic to its image in M''. Thus $P \in \operatorname{Ass} M''$ as required. □

The first exact sequences on which we shall use Lemma 3.6 are produced as follows:

Proposition 3.7. *If R is a Noetherian ring and M is a finitely generated R-module, then M has a filtration*

$$0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$$

with each $M_{i+1}/M_i \cong R/P_i$ for some prime ideal P_i.

Proof. If $M \neq 0$ then by Proposition 3.4, M has at least one associated prime, say P_1, so that there is a submodule $M_1 \cong R/P_1$. Applying this reasoning again to M/M_1, we produce M_2, and continue in this way. The process must come to an end because the submodules of M satisfy the ascending chain condition, and this means that some $M_n = M$, as required. □

Using Lemma 3.6 inductively, we see that the associated primes of M are among the primes P_i appearing in Proposition 3.7. This proves the finiteness statement of Theorem 3.1.

One might ask which modules M admit a filtration as in Proposition 3.7, where in addition every P_i is an associated prime of M. Such modules are called **clean**. For example, when R is a domain and M is torsion-free but not free, M is not clean, as the reader may verify. As of this writing I know of no interesting characterization of cleanliness—perhaps the reader will find one! Proposition 3.13 provides an interesting class of filtrations where the associated primes do split up nicely.

Conclusion of the Proof of Theorem 3.1. We first prove part c: If $P \in \operatorname{Ass} M$, then there is an inclusion $R/P \subset M$. Localizing, we get an injection $R[U^{-1}]/PR[U^{-1}] \subset M[U^{-1}]$. Thus if $PR[U^{-1}]$ is a prime ideal of $R[U^{-1}]$—that is, if $P \cap U = \emptyset$ so $PR[U^{-1}]$ is still a proper ideal—then $PR[U^{-1}] \in \operatorname{Ass} M[U^{-1}]$.

Conversely, suppose Q is a prime of $R[U^{-1}]$ that is associated to $M[U^{-1}]$. We may write $Q = PR[U^{-1}]$ with P a prime of R and $P \cap U = \emptyset$. There is an injection $\varphi : R[U^{-1}]/PR[U^{-1}] \to M[U^{-1}]$. Since P is finitely generated, we have

$$\operatorname{Hom}_{R[U^{-1}]}(R[U^{-1}]/PR[U^{-1}], M[U^{-1}]) = \operatorname{Hom}_R(R/P, M)[U^{-1}]$$

by Proposition 2.10, so we may write $\varphi = u^{-1}f$ for some $f \in \operatorname{Hom}_R(R/P, M)$ and $u \in U$. Since u is a nonzerodivisor on R/P, it follows that f is an injection, concluding the proof of part c.

It remains to show that if P is any prime minimal over ann M, then $P \in \operatorname{Ass} M$. By part c, we may localize and suppose that R is local with maximal ideal P. By Proposition 3.4 the set $\operatorname{Ass} M$ is nonempty, and since P is the only prime that contains ann M, it follows that $P \in \operatorname{Ass} M$. □

3.3 Primary Decomposition

To avoid endlessly repeating the hypotheses, we shall assume throughout the rest of this chapter that R is a Noetherian ring, and we shall assume that M is a finitely generated R-module.

As often happens, it is advantageous to work with modules instead of ideals, and we shall define primary decompositions for a submodule M' of a finitely generated module M: That is, we shall write M' as the intersection of certain submodules M_i that correspond to the prime powers above. These are defined as follows: A submodule N of a module M is **primary** if $\operatorname{Ass}(M/N)$ consists of just one prime; if $\operatorname{Ass}(M/N) = \{P\}$, we say that N is ***P*-primary**. Since this is really a condition on M/N, it is convenient to say that a module M is **coprimary** if 0 is a primary submodule—that is, if $\operatorname{Ass}(M)$ consists of just one prime ideal. From Lemma 3.6 we easily see that an intersection of P-primary submodules is P-primary.

Corollary 3.8. *Suppose that P is a prime ideal of a ring R, and $N_1, \ldots, N_t \subset M$ are R-modules. If each N_i is a P-primary submodule of M, then $\cap_i N_i$ is P-primary.*

Proof. By induction it suffices to do the case $t = 2$. By hypothesis M/N_1 and M/N_2 are P-coprimary. Lemma 3.6a shows that P is the only associated prime of $M/N_1 \oplus M/N_2$. Since $M/(N_1 \cap N_2)$ injects into $M/N_1 \oplus M/N_2$, Lemma 3.6b shows that $M/(N_1 \cap N_2)$ is also coprimary. □

The results of Theorem 3.1 lead to the following description of coprimary modules.

Proposition 3.9. *Let P be a prime ideal of R. The following statements are equivalent:*

a. *M is P-coprimary.*

b. *P is minimal over ann M, and every element not in P is a nonzerodivisor on M.*

c. *A power of P annihilates M, and every element not in P is a nonzerodivisor on M.*

Proof.

a $\Rightarrow$ b: Since P is the only associated prime of M, Theorem 3.1a shows that P is minimal over ann M, and Theorem 3.1b shows that every element not in P is a nonzerodivisor on M.

b $\Rightarrow$ c: Since the elements not in P are nonzerodivisors on M, it suffices to prove the statement after localizing at P, so we may assume that R is a local ring with maximal ideal P. Since P is minimal over ann M, it follows by Corollary 2.12 that P is the radical of ann M, so P is nilpotent modulo ann M.

c $\Rightarrow$ a: Since P is nilpotent modulo ann M, it is certainly minimal among primes containing ann M and is an associated prime of M by Theorem 3.1a. Since every element outside of P is a nonzerodivisor, every associated prime of M is contained in P by Theorem 3.1b. Thus P is the only associated prime of M.□

The most important case is the one where $M = R/I$ for some ideal I of R. In this setup, Proposition 3.9c shows that I is P-primary iff I contains a power of P, and for every $r, s \in R$, the conditions $rs \in I$ and $r \notin P$ imply $s \in I$. This is the classical definition.

It is often convenient to think of these definitions above in terms of localizations: Proposition 3.9b shows that M is P-coprimary iff P is minimal over the annihilator of M and M injects into M_P. In general, if M is any module and P is a minimal prime over the annihilator of M, then the submodule $M' \subset M$ defined by

$$M' = \ker(M \to M_P)$$

is P-primary because M/M' injects into $(M/M')_P = M_P$. In this situation, M' is called the ***P*-primary component** of 0 in M. Note that it depends only on M and on P.

Primary decomposition consists of writing an arbitrary submodule M' of M as the intersection of primary submodules.

Theorem 3.10. *Let R be a Noetherian ring, and let M be a finitely generated R-module. Any proper submodule M' of M is the intersection of primary submodules. Furthermore, if $P_1, \ldots, P_n$ are prime ideals and we write $M' = \bigcap_{i=1}^n M_i$ with M_i a P_i-primary submodule, then*

a. *Every associated prime of M/M' occurs among the P_i.*

b. *If the intersection is* ***irredundant*** *(meaning no M_i can be dropped), then the P_i are precisely the associated primes of M/M'.*

c. *If the intersection is* ***minimal****, in the sense that there is no such intersection with fewer terms, then each associated prime of M/M'*

is equal to P_i *for exactly one index* i. *In this case, if* P_i *is minimal over the annihilator of* M/M', *then* M_i *is the* P_i*-primary component of* M'.

d. *Minimal primary decompositions localize in the following sense: Suppose that* $M' = \bigcap_{i=1}^{n} M_i$ *is a minimal primary decomposition. If* U *is any multiplicatively closed set of* R, *and* $P_1, \ldots, P_t$ *are the primes among the* P_i *that do not meet* U, *then*

$$M'[U^{-1}] = \bigcap_{i=1}^{t} M_i[U^{-1}]$$

is a minimal primary decomposition over $R[U^{-1}]$.

Proof. We first prove the existence of a slightly finer but less canonical decomposition. We shall say that a submodule $N \subset M$ is **irreducible** if N is not the intersection of two strictly larger submodules. We first claim—and this is Emmy Noether's fundamental observation—that every submodule of M can be expressed as the intersection of irreducible submodules. Otherwise, by the ascending chain condition on submodules of M, we could choose a submodule $N \subset M$ maximal among those submodules that are not the intersection of irreducible submodules. In particular, N itself is not irreducible, so it is the intersection of two strictly larger submodules N_1 and N_2. By the maximality of N, both the N_i are intersections of irreducible submodules, and it follows that N is too. The contradiction proves our claim and shows that there is an **irreducible decomposition** $M' = \cap_i M_i$ with each M_i irreducible.

We next show that every irreducible decomposition is a primary decomposition. That is, we show that any irreducible submodule $N \subset M$ is primary, or equivalently that M/N is coprimary. Otherwise, M/N would have at least two associated primes, P and Q say, so it would contain a submodule isomorphic to R/P and another isomorphic to R/Q. The annihilator of every nonzero element of R/P is P, and similarly for Q, so these two submodules of M/N can only meet in 0. Thus 0 is reducible. Taking preimages of these submodules in M, we see that N is reducible: a contradiction. This proves that M/N is coprimary, and thus that irreducible decompositions are primary decompositions.

Statements a through d are really statements about M/M'. To simplify the notation, we begin by factoring out M', and we assume henceforward that $M' = 0$.

a. Now suppose that $0 = \bigcap_i M_i$ is a primary decomposition. Note that $M \subset \oplus M/M_i$, so by Lemma 3.6 every prime in $\operatorname{Ass} M$ occurs among the primes P_i. This proves assertion a.

b. Next suppose that the given decomposition is irredundant, so that for each j, $\bigcap_{i \neq j} M_i \neq 0$. Note that because $M_j \cap \bigcap_{i \neq j} M_i = 0$, we

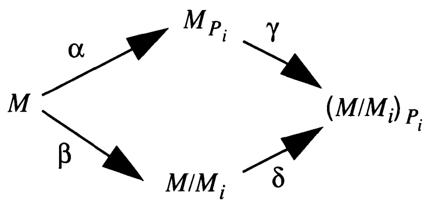
have

$$\bigcap_{i\neq j} M_i = \left(\bigcap_{i\neq j} M_i\right) \Big/ \left(M_j \cap \bigcap_{i\neq j} M_i\right)$$
$$\cong \left(\bigcap_{i\neq j} M_i + M_j\right) /M_j \subset M/M_j.$$

As this module is P_j-coprimary, so is $\bigcap_{i\neq j} M_i$. By Lemma 3.6, P_j is an associated prime of M. Together with a, this proves part b.

c. Finally, suppose that the given decomposition is minimal. By Corollary 3.8 the intersection of P-primary submodules is P-primary, so minimality implies that the primes P_i are distinct. With part b, this proves the first statement of c.

For the last statement, suppose that P_i is minimal over the annihilator of M. We must show that M_i is the kernel of the localization map $\alpha : M \to M_{P_i}$. Consider the commutative diagram



where β is the projection map, δ is the localization map, and γ is the projection of M_{P_i} to $M_{P_i}/M_{iP_i} = (M/M_i)_{P_i}$. The kernel of β is M_i. To show that the kernel of α is also M_i, it suffices to show that both γ and δ are monomorphisms. Since M_i is P_i-primary, this is immediate for δ.

Since $\cap_j M_j = 0$, the natural map $\varphi : M \to \oplus M/M_j$ is a monomorphism. By Proposition 2.5, localization preserves monomorphisms, so $\varphi_{P_i} : M_{P_i} \to \oplus(M/M_j)_{P_i}$ is a monomorphism. The map γ is the ith component of φ_{P_i}. Because P_i is minimal over the annihilator of M, we know that P_j is not contained in P_i for $j \neq i$. Since M/M_j is P_j-coprimary, we have $(M/M_j)_{P_i} = 0$ for $j \neq i$, so the jth component of φ_{P_i} vanishes, and we see that γ is a monomorphism as required.

d. If $U \cap P_i = \emptyset$, then $P_i[U^{-1}]$ is a prime ideal of $R[U^{-1}]$, and by Theorem 3.1c, $M_i[U^{-1}]$ is $P_i[U^{-1}]$-primary. If $U \cap P_i \neq \emptyset$ then we see from Proposition 3.9c that $M_i[U^{-1}] = M[U^{-1}]$. Thus

$$0 = \bigcap_{i=1}^{t} M_i[U^{-1}]$$

is a primary decomposition. To see that it is minimal, it suffices by part b to show that the associated primes of $M[U^{-1}]$ are the associated primes of M that do not meet U, and this also follows from Theorem 3.1c. □

In Exercise A3.6 we present a different view of primary decomposition: It is the reflection, in M, of the fact that the injective envelope of M decomposes in a nice way. This point of view also explains the meaning of the irreducible decompositions defined in the preceding proof.

3.4 Primary Decomposition and Factoriality

It is easy to express the relationship between primary decomposition and unique factorization in the classical sense.

Proposition 3.11. *Let R be a Noetherian domain.*

a. If $f \in R$ and $f = u\Pi p_i^{e_i}$, in such a way that u is a unit of R, the p_i are primes generating distinct ideals (p_i), and each e_i is a positive integer, then $(f) = \cap(p_i^{e_i})$ is the minimal primary decomposition of (f).

b. R is factorial iff every prime ideal minimal over a principal ideal is itself principal.

Proof.

a. First we show that $(p_i^{e_i})$ is a (p_i)-primary ideal. If Q is an associated prime of $(p_i^{e_i})$, then since Q contains a power of p we have $Q \supset (p_i)$. If q is any element of Q, then q annihilates some element of $R/(p_i^{e_i})$; that is, for some $f \notin (p_i^{e_i})$ we have $qf = p_i^{e_i}g$. Since $p_i^{e_i}$ divides qf but not f, and since p_i is prime, we see that p_i divides q. This shows $Q \subset (p_i)$ as required.

Clearly, we have $(f) \subset \cap(p_i^{e_i})$; we wish to show equality. By induction on the number of primes p_i involved, it suffices to show that if g is not divisible by a prime p, then $(g) \cap (p^e) = (gp^e)$. But if $hg \in (p^e)$, then since p does not divide g and p is prime, p must divide h, and $(h/p)g \in (p^{e-1})$. Repeating this argument, we eventually see that p^e divides h, so $hg \in (gp^e)$.

We now see that $(f) = \cap(p_i^{e_i})$ is a primary decomposition. Thus every prime of $\operatorname{Ass} R/(f)$ is one of the (p_i). Each (p_i) is contained in an associated prime of (f) because p_i is a zerodivisor modulo (f): for p_i divides f and $p_i(f/p_i) \in (f)$. Thus, the given primary decomposition is minimal.

b. Suppose R is factorial. If $f = u\Pi p_i^{e_i}$ is the prime factorization of an element, then by part a the associated primes of (f), and thus in

particular the minimal primes of R that contain f, are the principal primes (p_i).

Conversely, suppose that every prime ideal minimal over a principal ideal is itself principal. To prove that R is factorial, the argument given in Section 0.2 shows that since R is Noetherian it is enough to check that any irreducible element $f \in R$ is prime. But if P is a prime minimal over (f), then by hypothesis we may write $P = (p)$ for some $p \in R$, and $f \in P$ becomes $f = rp$ for some $r \in R$. Since f is irreducible, r must be a unit, so $(f) = (p) = P$ is prime. □

We shall sharpen this result a little in Corollary 10.6.

3.5 Primary Decomposition in the Graded Case

If R is a graded Noetherian ring and M is a finitely generated graded R-module, then the associated primes of M are homogeneous, a primary decomposition of 0 in M can be made in terms of homogeneous modules, and M has a filtration as in Proposition 3.7 where the M_i and P_i are homogeneous. The proofs of these things involve only one new idea, given in Proposition 3.12, and we leave the details to the reader. We state the proposition here for ordinary graded rings $R = R_0 \oplus R_1 \oplus \cdots$, but in fact it holds (with the same proof!) for **Z**-graded rings and modules, and much more generally. See Exercise 3.5.

Proposition 3.12. *Suppose that $R = R_0 \oplus R_1 \oplus \cdots$ is a graded ring, and M is a graded R-module. Let $m \in M$ be any element, and set $P = \operatorname{ann} m \subset R$. If P is prime, then P is homogeneous and P is the annihilator of a homogeneous element.*

Proof. Any $f \in R$ may be expressed in a unique way as a sum $f = \sum_{i=1}^{s} f_i$, where each f_i is nonzero and homogeneous of some degree d_i, and $d_1 < \cdots < d_s$. We may prove that P is homogeneous by showing that if $f \in P$ then $f_i \in P$ for each i. By induction on s it suffices to show that $f_1 \in P$. Thus we suppose that $fm = 0$ and we wish to prove that $f_1 m = 0$.

We may also write $m = \sum_{i=1}^{t} m_i$ in a unique way so that each m_i is nonzero and homogeneous of some degree e_i, and $e_1 < \cdots < e_t$. We do induction on the number of terms t. Since $fm = f_1 m_1 +$ (terms of higher degree), we see that $f_1 m_1 = 0$. Thus, if $t = 1$ we are done. Suppose $t > 1$ and that the result has been proven for all smaller values of t.

The element $f_1 m = \sum_{i=2}^{t} f_1 m_i$ is a sum of fewer homogeneous terms than is m. Set $I = \operatorname{ann} f_1 m$. Note that $P \subset I$. If $P = I$ then P is homogeneous by the induction, and we are done. Otherwise, we may choose an element $g \in I$ such that $g \notin P$. We have $g f_1 m = 0$, so $g f_1 \in \operatorname{ann} m = P$. Since

$g \notin P$, and P is prime, we have $f_1 \in P$ as claimed, proving that P is homogeneous.

From the fact that P is homogeneous it follows that $Pm_i = 0$ for each i. Since $P = \operatorname{ann} m \supset \cap_i(\operatorname{ann} m_i) \supset P$, we see that $P = \cap_i(\operatorname{ann} m_i) \supset \Pi_i(\operatorname{ann} m_i)$. Since P is prime, we have $P \supset \operatorname{ann} m_i$ for some i, whence $P = \operatorname{ann} m_i$, and we are done. □

3.6 Extracting Information from Primary Decomposition

We maintain the assumptions that R is a Noetherian ring, and we shall assume that M is a finitely generated R-module.

We have already seen that if $0 = \cap_i M_i$ is the minimal primary decomposition, then the M_i corresponding to minimal primes of Ass M are uniquely determined by M, and thus might be expected to shed some light on the structure of M, whereas the M_i corresponding to embedded primes generally are not uniquely determined (we shall analyze this phenomenon in a moment). The same mechanism that leads to the uniqueness of the M_i corresponding to the minimal primes carries us a little further and shows that certain intersections of primary components are well defined. It turns out that these intersections correspond to the sets of associated primes not containing a given ideal—that is, to the closed subsets of Spec A in the Zariski topology introduced in Chapter 1.

To express the intersections above, we shall make a definition: For any ideal $I \subset R$, we set

$$\mathbf{H^0_I(M)} = \{m \in M \mid I^n m = 0 \quad \text{for } n \gg 0\}$$

the set of elements annihilated by some power of I. The notation comes from local cohomology; see Appendix A4, in which functors $H^i_I(M)$ are defined for all i. (Pursuing the analogy with sheaf theory from which local cohomology arises, some authors write $\Gamma_I(M)$ for what we have called $H^0_I(M)$.)

The set $H^0_I(M)$ is easily seen to be a submodule of M. It actually depends only on the radical of I, in the sense that $H^0_I(M) = H^0_J(M)$ if $\operatorname{rad}(I) = \operatorname{rad}(J)$.

Proposition 3.13. *Let I be an ideal of R, and let*

$$A = \{P \in \operatorname{Ass} M | P \supset I\}.$$

a. *Let $0 = \cap_i M_i$ be a primary decomposition of $0 \subset M$, and suppose M_i is P_i-primary. The submodule $H^0_I(M)$ is the intersection of those M_i such that $P_i \notin A$. In particular, this intersection is independent of the primary decomposition chosen.*

b. *There is an element* $f \in I$ *such that* $P \in A$ *iff* $P \in \operatorname{Ass} M$ *and* $f \in P$. *For any such* f *we have*

$$H_I^0(M) = \ker(M \to M[f^{-1}]).$$

c. *We have* $\operatorname{Ass} H_I^0(M) = A$, *and* $\operatorname{Ass} M/H_I^0(M) = (\operatorname{Ass} M) - A$. *These properties characterize* $H_I^0(M)$ *uniquely.*

Proof.

a. We may write $H_I^0(M) = (0 :_M I^\infty) := \cup_n (0 :_M I^n)$, where $(0 :_M I^n) = \{m \in M | I^n m = 0\}$. Using the given primary decomposition, we get

$$H_I^0(M) = \left(\left(\bigcap_i M_i \right) :_M I^\infty \right) = \bigcap_i (M_i :_M I^\infty) .$$

A power of P_i annihilates M/M_i, so if $P_i \supset I$ then $(M_i :_M I^\infty) = M$, and we may drop this component from the intersection. On the other hand, if $P_i \not\supset I$ then I contains a nonzerodivisor on M/M_i, so $(M_i :_M I^\infty) = M_i$. The desired formula for $H_I^0(M)$ follows.

b. By prime avoidance we may choose $f \in I$ not in any of the finitely many primes $Q \in (\operatorname{Ass} M) - A$. Set $N = \ker(M \to M[f^{-1}])$. By Proposition 2.1 we have $N = (0 :_M f^\infty)$. By the argument of part a, applied to the ideal (f) in place of I, this is the intersection of those M_i such that $P \not\supset f$, the same as $H_I^0(M)$.

c. By part a, the primary decomposition of $H_I^0(M)$ in M is

$$H_I^0(M) = \bigcap_{i \text{ such that } P_i \notin A} M_i.$$

If we choose the primary decomposition $0 = \cap_i M_i$ to be irredundant, then we get an irredundant primary decomposition of $H_I^0(M)$, and it follows from Theorem 3.10 that $\operatorname{Ass} M/H_I^0(M) = (\operatorname{Ass} M) - A$. Further, by Lemma 3.6b we see that Ass $H_I^0(M)$ is a subset of primes of Ass M that contains A. Since every element of $H_I^0(M)$ is annihilated by a power of I, it follows that the primes of Ass $H_I^0(M)$ all contain I. Thus Ass $H_I^0(M) = A$.

Conversely, let N be any submodule of M such that Ass $N = A$ and $\operatorname{Ass} M/N = \operatorname{Ass} M - A$. If we choose f as in part b, then a power of f annihilates N and f is a nonzerodivisor on M/N. If follows that $N = \ker(M \to M[f^{-1}])$, so $N = H_I^0(M)$ by part b. □

The mechanism of part b could be applied with any localization, but it does not yield any submodules other than the $H_I^0(M)$. See Exercise 3.12.

A typical application of part a of the proposition is to show that the intersection of all primary components corresponding to primes of dimension greater than or equal to some number d is well defined. (See Chapter 9 for the definition of dimension.)

The most interesting case of Proposition 3.13 occurs when the ideal I is a prime P. The module $H^0_P(M)_P \subset M_P$ is then the unique largest submodule of finite length. Its length is called the **multiplicity of P in M**. We see from the proposition (or directly from Theorem 3.1) that P is associated to M iff the multiplicity of P in M is nonzero. In general, one may think of the multiplicity as measuring "how associated" P is to M.

Somewhat surprisingly, there seems no general way to extract "invariant" information about M from a primary decomposition that is not covered by Proposition 3.13 (but in some special circumstances there is—see, for example, Exercise 3.11). This has led some people to the view that one should ignore primary decomposition entirely; localization and the set of associated primes together are sufficient for many purposes.

3.7 Why Primary Decomposition Is Not Unique

We take a moment to explain why the terms in a primary decomposition corresponding to embedded primes are not unique, and to explore some related ideas. Assume for simplicity that R is a local Noetherian ring, and that the finitely generated module M has two associated primes, a minimal prime Q and the maximal ideal P itself. If we write a minimal primary decomposition $0 = M' \cap M''$, where M' is Q-primary and M'' is P-primary, then by Theorem 3.10c, $M' = \ker(M \to M_Q)$ is uniquely determined. However, as the reader may easily check, M'' may be taken to be any submodule such that

a. For some integer d, $M'' \supset P^d M$.

b. $M'' \cap M' = 0$.

In particular, we could simply take $M'' = P^d M$ for any sufficiently large d.

One may try to avoid the problem by taking M'' maximal satisfying properties a, b. However, uniqueness is prevented even then, essentially by the fact that the complement of a vector space is not unique. For example, let k be a field and let $R = k[x]_{(x)}$ be a localization of the polynomial ring in one variable. Let $M = R \oplus R/(x)$, and let e be a generator for the second summand. With notation as above, $Q = 0$, $P = (x)$, and $M' = Re$, the second summand. Here M'' may be any nonzero submodule meeting Re in 0. The maximal choices for M'' are precisely the complements of the second summand, Re; these are the modules generated by elements of the form $(1, ue)$, with $u \in k$. Since any two such elements are carried into one another

by an automorphism of M, there is no distinguished choice for M''. (Some more examples are given in Exercise 3.10.)

In situations where "nice" subspaces have distinguished complements (for example, in the presence of a suitable group action) there are sometimes distinguished primary decompositions, however. See Exercise 3.11.

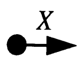
3.8 Geometric Interpretation of Primary Decomposition

If k is an algebraically closed field and $I \subset S = k[x_1, \ldots, x_r]$ is an ideal, we can hope to "see" some of the meaning of a primary decomposition of I. Let $I = \cap_j I_j$ be a minimal primary decomposition. It follows of course that $Z(I) = \cup_j Z(I_j)$. If I is a radical ideal, then each of the I_j is a prime ideal minimal over I, and the primary decomposition simply expresses the algebraic set $Z(I)$ as the union of the irreducible algebraic sets (algebraic varieties) $Z(I_j)$. But in more general cases the algebra suggests more. What we shall do here informally is formalized in the theory of schemes; see, for example, Eisenbud and Harris [1992] for an expository treatment in the spirit of this text, and Hartshorne [1977, Chapter 2] for more technical detail.

Let us begin with the case of an ideal $I \subset S = k[x, y]$ that is primary to the maximal ideal (x, y) so that $Z(I)$ is the origin in the affine plane. For example, what geometric object X should be associated with the primary ideal (x^2, y)? The idea is that X should be that geometric object that determines the coordinate ring S/I. If

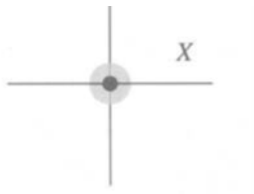
$$f = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 + a_6 x^3 + \cdots$$

is a polynomial, then from the class of f modulo (x^2, y) we can read off the scalars $a_0 = f(0,0)$ and $a_1 = \partial f/\partial x(0,0)$. That is, if we restrict a function to X, then we "see" the value of the function at the origin—so the point (0, 0) should be "in" X—and the value of the first derivative of f in the horizontal direction. There is a standard geometric object of this kind: It is the origin plus the horizontal tangent vector at the origin!



$I = (x^2, y)$ corresponds to X, a point with tangent vector.

In a similar way, if we take $I = (x^2, xy, y^2)$, then the class of f modulo I reveals the value of f at 0 and the value of the first derivative of f in any direction. Thus, it is natural to think of the corresponding X as the whole first-order infinitesimal neighborhood of the origin in the plane.



$I = (x^2, xy, y^2)$ corresponds to X, the first-order infinitesimal neighborhood of (0,0).

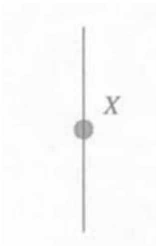
If we replace I by, for example, the nth power $(x, y)^n$, then all the derivatives of f up to order $n - 1$ are visible modulo I, so the corresponding geometric object X is the $(n - 1)$st infinitesimal neighborhood.

Similar considerations are suggestive in higher dimensional cases, too. For example, the ideal $(x) \subset k[x, y]$ corresponds to $Z((x))$, the vertical line in the plane, while modulo (x^2) one can see the values of a function $f(x, y)$ at every point on the vertical line together with the values of its first derivatives in the horizontal direction at any point of the line. Thus (x^2) corresponds to the vertical line with all the horizontal tangent vectors at points of the line—that is, the first-order neighborhood of the vertical line:



$I = (x^2)$ corresponds to X, the first-order infinitesimal neighborhood of the vertical line.

From these ideas it is easy to see how to interpret more-or-less arbitrary primary decompositions. For example, $I = (x) \cap (x^2, xy, y^2)$ corresponds to the vertical line together with the first-order neighborhood at the origin.



$I = (x) \cap (x^2, xy, y^2)$ corresponds to X, the vertical line plus the first-order infinitesimal neighborhood of (0, 0).

Here the primary decomposition is not unique, and we could also write $I = (x^2, xy) = (x) \cap (x^2, y)$, corresponding to the fact that the only information about a function f that is available on the first-order infinitesimal neighborhood of the origin but not on the vertical line is the derivative of the function in the horizontal direction.

3.9 Symbolic Powers and Functions Vanishing to High Order

If P is a maximal ideal of R and I is any proper ideal containing a power of P, then I is P-primary: For in this case P is the only prime containing the annihilator I of R/I, so Theorem 3.1a shows that $\operatorname{Ass} R/I = \{P\}$. This generalizes the fact that any power of a prime in the integers is primary.

In particular, the powers of a maximal ideal are all primary. One would be tempted to hope that a power of any prime ideal P would be P-primary, but this is not the case. In general, the P-primary component of the nth power of P is called the **nth symbolic power** of P, and is written $P^{\langle n \rangle}$. In the geometric case, the symbolic powers of P have a nice geometric description as follows, due to Zariski and Nagata.

Suppose that k is an algebraically closed field of characteristic 0 and $S = k[x_1, \ldots, x_r]$ is a polynomial ring. Let X be the variety corresponding to the prime ideal $P \subset S$, so that P is the set of all polynomials vanishing on X. For $n \geq 1$, let

$$P^{\langle n \rangle} = \{f \in S | f \text{ vanishes to order } \geq n \text{ at every point of } X\}.$$

The condition that f vanishes to order n at a point $x \in \mathbf{A}^r$ means that if $\mathfrak{m}_x$ is the maximal ideal of S consisting of functions vanishing at x, then $f \in \mathfrak{m}_x^n$; equivalently, the Taylor expansion of f around x begins with terms of order greater than or equal to n. Thus we may also write

$$P^{\langle n \rangle} = \bigcap_{x \in X} \mathfrak{m}_x^n.$$

If the characteristic of k is 0, then $P^{\langle n \rangle}$ can be defined in another way as well: It is the set of polynomials that vanish together with their partial

derivatives of orders less than n at all the points of X. (In characteristic p, this is a weaker condition, and not so interesting: the derivatives of order $\geq p$ of the function x_1^m are identically 0.)

Theorem 3.14 (Zariski, Nagata). *Suppose that k is an algebraically closed field and S is a polynomial ring over k. If P is a prime ideal of S, then $P^{\langle n \rangle} = P^{(n)}$, the nth symbolic power.*

Theorem 3.14 is true (with suitable interpretation) in a much broader setting. See Eisenbud and Hochster [1979] for history and details.

Partial Proof. We shall prove in characteristic 0 that $P^{\langle n \rangle}$ is P-primary and contains P^n. It follows that $P^{\langle n \rangle}$ contains $P^{(n)}$. We only sketch the opposite inclusion; for a full proof see Eisenbud and Hochster [1979] and its references. It is obvious that $P^{\langle n \rangle}$ is an ideal and that $P^{\langle n \rangle} \supset P^n$. To show that $P^{\langle n \rangle}$ is P-primary, we must show that if $r \notin P$, but $rs \in P^{\langle n \rangle}$, then $s \in P^{\langle n \rangle}$.

If $\mathfrak{m}$ is a maximal ideal of S containing P such that $r \notin \mathfrak{m}$, then since $rs \in \mathfrak{m}^n$ and $\mathfrak{m}^n$ is $\mathfrak{m}$-primary, we must have $s \in \mathfrak{m}^n$. It follows that the derivatives of order less than n of s all vanish on the set $Y = \{x \in X | r(x) \neq 0\}$. Let g be such a derivative. Since rg vanishes at every point of X, we have $rg \in P$ by the Nullstellensatz. Since $r \notin P$ by hypothesis, it follows that $g \in P$. Under the hypothesis that k has characteristic 0 we deduce that s vanishes to order $\geq n$ on X, proving that $P^{\langle n \rangle}$ is P-primary.

Here is the idea of the proof that $P^{\langle n \rangle} \subset P^{(n)}$: Since $P^{(n)}$ is P-primary, it is enough to show that $(P^{\langle n \rangle})[U^{-1}] \subset (P^{(n)})[U^{-1}]$ for some multiplicatively closed set U not meeting P. We shall later show that there exists an element $u \notin P$ such that for any point $x \in X$ with $u(x) \neq 0$, with corresponding maximal ideal $\mathfrak{m} = \mathfrak{m}_x$, there is a set of generators $y_1, \ldots, y_r$ of $\mathfrak{m}_\mathfrak{m}$ such that $P_\mathfrak{m}$ is generated by a subset of the y_i. Under these circumstances the y_i act like a set of "variables" (see Corollary 10.14 and Exercise 17.13).

To see how the argument should go, we shift to the simpler case where P is generated by a subset of variables: $P = (y_1, \ldots, y_c) \subset k[y_1, \ldots, y_r]$. The polynomials $f(y_1, \ldots, y_r)$, all of whose derivatives of order less than n are in $(y_1, \ldots, y_c)$, are precisely the polynomials whose terms are all of degree at least n in $y_1, \ldots, y_c$—that is, they are the polynomials in the nth power of P, and $P^{\langle n \rangle} = P^n$. The nth power of $(y_1, \ldots, y_c)$ is also primary by Exercise 3.6, so $P^n = P^{(n)}$. The analogous statements are also true in the original case. In particular, after inverting u we have $P^{\langle n \rangle} = P^n = P^{(n)}$. □

3.9.1 *A Determinantal Example*

These ideas suggest an explicit example of a prime whose square is not equal to its symbolic square (and we shall check the example directly). A good general reference for the material that follows is the book of Bruns and Vetter [1988], and the example we shall give is very close to the paper

of DeConcini, Eisenbud, and Procesi [1982]; in particular, all the unproved assertions encountered below are proved in these sources.

Consider the polynomial ring in 9 variables $S = k[\{x_{ij}\}_{1\leq i,j\leq 3}]$ and the generic 3×3 matrix $G = (x_{ij})$ over S. Let P be the radical of the ideal $I_2(G)$ generated by the 2×2 minors of G. The algebraic set X defined by $I_2(G)$ in the set $M_3 = \mathbf{A}^9$ of all 3×3 matrices is the set of 3×3 matrices of rank ≤ 1. This set is irreducible, so that P is prime, as the following very typical geometric argument shows.

First, the algebraic set

$$Y := GL(3,k) = \{(g,y) \subset \mathbf{A}^9 \times \mathbf{A}^1 | g \text{ a } 3 \times 3 \text{ matrix and } (\det g)y = 1\}$$

is irreducible because the corresponding ring is $k[\{x_{ij}\}_{1\leq i,j\leq 3}][(\det g)^{-1}]$, a localization of the polynomial ring in 9 variables. The same is true of the algebraic set $Y \times Y \subset \mathbf{A}^{20}$; its ring is a localization of the ring of polynomials in 18 variables. Let $M_3 = \mathbf{A}^9$ be the set of 3×3 matrices over k. Choose any matrix m of rank exactly 1, and consider the morphism $Y \times Y \to M_3$ defined by $(g,h) \mapsto gmh$. Because any two nonzero matrices of rank 1 differ only by a change of basis in source and target, the image of φ is exactly X. If $X = X_1 \cup X_2$, with X_1 and X_2 algebraic subsets of X, then $\varphi^{-1}(X_1) \cup \varphi^{-1}(X_2) = Y \times Y$. Since $Y \times Y$ is irreducible, either $\varphi^{-1}(X_1) = Y \times Y$ or $\varphi^{-1}(X_2) = Y \times Y$, and thus $X_1 = X$ or $X_2 = X$, showing that X is irreducible, too.

It is obvious that no linear form vanishes on all rank-1 matrices, so P contains no linear form. In fact, $I_2(G)$ is prime, so $P = I_2(G)$ is the prime ideal of functions vanishing on the set of rank-1 matrices, but we shall not need this here.

Let $g = \det G$, the determinant of G. We claim that $g \in P^{(2)}$. Since P contains no linear forms, P^2 is generated by forms of degree ≥ 4 and g is of degree 3, so this will show that $P^2 \neq P^{(2)}$.

Checking Theorem 3.14 against this example, we note that the partial derivatives of g with respect to the variables x_{ij} are 2×2 minors of G, so $g \in P^{\langle 2\rangle}$. If k has characteristic 0, then Theorem 3.14 applies to show that $g \in P^{(2)}$ as claimed.

We now give a direct proof. We must show that g becomes an element of P^2 after we localize at P. Now $x_{11} \notin P$, so it suffices to show that $x_{11}g \in I_2(G)^2$. This is easy to check: After multiplying the second and third columns of G by x_{11}, which changes the determinant to $x_{11}^2 g$, we may add multiples of the first column to the two other columns (not changing the determinant) to make the 1, 2 and 1, 3 elements of the matrix 0, as in Figure 3.1:

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \mapsto \begin{pmatrix} x_{11} & x_{11}x_{12} & x_{11}x_{13} \\ x_{21} & x_{11}x_{22} & x_{11}x_{23} \\ x_{31} & x_{11}x_{32} & x_{11}x_{33} \end{pmatrix} \mapsto$$

$$\begin{pmatrix} x_{11} & 0 & 0 \\ x_{21} & x_{11}x_{22} - x_{12}x_{21} & x_{11}x_{23} - x_{13}x_{21} \\ x_{31} & x_{11}x_{32} - x_{12}x_{31} & x_{11}x_{33} - x_{13}x_{31} \end{pmatrix}.$$

FIGURE 3.1.

Thus the determinant $x_{11}^2 g$ is the product of x_{11} and the determinant of the lower 2×2 submatrix

$$G' = \begin{pmatrix} x_{11}x_{22} - x_{12}x_{21} & x_{11}x_{23} - x_{13}x_{21} \\ x_{11}x_{32} - x_{12}x_{31} & x_{11}x_{33} - x_{13}x_{31} \end{pmatrix},$$

so that $x_{11}g = \det G'$. Since the entries of G' are 2×2 minors of the original matrix, $\det G' \in I_2(G)^2$, and thus $g \in P^{(2)}$.

In fact, it is known that $P^{(2)} = (P^2, g)$, and that a primary decomposition of P^2 is $P^2 = P^{(2)} \cap \mathfrak{m}^4$, where $\mathfrak{m}$ is the ideal generated by all the x_{ij}.

Here is a geometric proof that g vanishes to order ≥ 2 at any point $a \in X$. Since we are in characteristic 0, it suffices to show that the partial derivative $\partial g/\partial x_{ij}$ vanishes at a for every i, j. If we write e_{ij} for the matrix which has all its entries equal to 0 except for the i, j entry, and whose i, j entry is 1, then $\partial g/\partial x_{ij} = dg(a + te_{ij})/dt$, where t is a new variable. But since both a and e_{ij} have rank 1, every matrix of the form $a + te_{ij}$ has rank ≤ 2. Thus g vanishes identically on matrices of the form $a + te_{ij}$, and we see that the derivative is 0 as required.

More generally, we might ask for the primary decomposition of any power of any "determinantal" ideal. To be specific, if $G = (x_{ij})$ is the "generic" $p \times q$ matrix over the ring $S = k[\{x_{ij}\}_{1 \leq i \leq p, 1 \leq j \leq q}]$ then for each n the ideal P_n generated by the $n \times n$ minors of G is prime. If $1 < n < \min(p, q)$, then the powers of P_n are not primary; however, the symbolic powers of P_n are known—they are generated by certain products of minors of various orders—and a primary decomposition of the powers has the form

$$(*) \qquad P_n^m = P_n^{(m)} \cap P_{n-1}^{(2m)} \cap \cdots \cap P_1^{(nm)}.$$

The decomposition $(*)$ can be made minimal by taking only the first $\alpha(m, n)$ terms for a certain function $\alpha(m, n)$—see DeConcini—Eisenbud and Procesi [1982] for a precise statement, proof, and history of these matters.

3.10 Exercises

Exercise 3.1: Let $R = \mathbf{Z}$, the ring of integers. Identify the associated primes of a finitely generated abelian group ($\mathbf{Z}$-module) in terms of the

usual structure theory of finitely generated abelian groups.

Exercise 3.2: If $M' = M_1 \cap M_2$ are submodules of a module M, show that $\operatorname{Ass} M/M' \subset \operatorname{Ass} M/M_1 \cup \operatorname{Ass} M/M_2$.

Exercise 3.3:* If R is Noetherian and M and N are finitely generated R-modules, show that

$$\operatorname{Ass} \operatorname{Hom}_R(M, N) = \operatorname{Supp} M \cap \operatorname{Ass} N,$$

where Supp M is the set of all primes containing the annihilator of M. (Hint: Show that it suffices to assume R is local and prove that the maximal ideal is in the set on the left-hand side iff it is in the set on the right-hand side. You will need to use Nakayama's lemma, Corollary 4.8.) Taking $M = R/I$, and setting $(0 :_N I) = \{n \in N | In = 0\}$, show that $\operatorname{Hom}_R(M, N) = (0 :_N I)$, and thus

$$\operatorname{Ass}(0 :_N I) = \operatorname{Ass} N \cap \{P \subset R | P \text{ is a prime ideal and } I \subset P\}.$$

Exercise 3.4 (Gauss' Lemma):* Let R be any ring, and set $S = R[x_1, \ldots, x_r]$, the polynomial ring in r variables. If $f \in S$ is a polynomial, write Content(f) for the ideal of R generated by the coefficients of f.

a. If f, $g \in S$ then

$$\operatorname{Content}(fg) \subset \operatorname{Content}(f) \operatorname{Content}(g) \subset \operatorname{rad}(\operatorname{Content}(fg)).$$

Deduce that if Content(f) contains a nonzerodivisor of R, then f is a nonzerodivisor of S.

b. If R is Noetherian and f is a nonzerodivisor of S, show conversely that Content(f) contains a nonzerodivisor of R.

c. We say that f is **primitive** if Content$(f) = (1)$. Gauss proved, in the case $R = \mathbf{Z}$ and $r = 1$, that the product of primitive polynomials is primitive, essentially to prove that if a primitive polynomial is irreducible in $\mathbf{Z}[x]$ then it is irreducible in $\mathbf{Q}[x]$. Prove that if R is a factorial domain with quotient field K, and if f is irreducible in $R[x]$, then f is irreducible in $K[x]$. Then show that $R[x]$ is again factorial. If R is an arbitrary factorial domain, then the natural analogue of a primitive polynomial is a polynomial f such that Content(f) is not contained in any principal ideal. Use part a to show that if R is a factorial domain then $R[x]$ is a factorial too. Deduce that if K is the quotient field of R and a polynomial $f \in R[x]$ is irreducible in $K[x]$ then f is irreducible in $R[x]$.

General Graded Primary Decomposition

Exercise 3.5: Let Γ be an abelian monoid (that is, a set with a commutative associative addition operation possessing an identity element 0), and let $R = \oplus_{\gamma \in \Gamma} R_\gamma$ be a ring graded by Γ, in the sense that each R_γ is an

abelian group and $R_\gamma R_{\gamma'} \subset R_{\gamma+\gamma'}$. We say that Γ acts on a set Λ if we are given a map $\Gamma \times \Lambda \to \Lambda$, denoted $(\gamma, \lambda) \mapsto \gamma + \lambda$ and the associative law $\gamma + (\gamma' + \lambda) = (\gamma + \gamma') + \lambda$ holds. We say that Γ acts freely on Λ if $\gamma + \lambda = \lambda$ only when $\gamma = 0$. If M is an R-module, we say that M is graded by Λ if $M = \oplus_{\lambda \in \Lambda} M_\lambda$ as abelian groups and $R_\gamma M_\lambda \subset M_{\gamma+\lambda}$ for any $\gamma \in \Gamma$, $\lambda \in \Lambda$. An element of R is called homogeneous if it belongs to one of the R_γ, and similarly for M. Every element of R or M can be written as a sum of nonzero homogeneous elements in a unique way; these are called its homogeneous components. An ideal $I \subset R$ is called homogeneous if it can be generated by homogeneous elements. Show that I is homogeneous iff I contains the homogeneous components of each of its elements.

If Γ and Λ are totally ordered then we say that the action of Γ on Λ is **compatible with** the order if $\gamma \leq \gamma'$ and $\lambda \leq \lambda'$ together imply $\gamma + \lambda \leq \gamma' + \lambda'$. We say that Γ acts **freely** if, under these circumstances, $\gamma + \lambda = \gamma' + \lambda'$ implies $\gamma = \gamma'$ and $\lambda = \lambda'$, and also that $\gamma + \lambda \neq \lambda$ unless $\gamma = 0$.

a. Suppose that Γ is a totally ordered abelian monoid and R is a ring graded by Γ. Suppose also that M is an R-module graded by a totally ordered set Λ on which Γ acts freely in a way compatible with the orders. If $P \subset R$ is a prime ideal that is the annihilator of an element of M, adapt the argument of Proposition 3.12 to show that P is homogeneous and that P is in fact the annihilator of a homogeneous element of M.

b. Suppose that R is a Noetherian ring, M is a finitely generated R-module, and R and M are graded as in part b. Show that Ass M consists of homogeneous prime ideals. Show that $0 \subset M$ has a primary decomposition into homogeneous submodules. Show that in the filtration of Proposition 3.7 the M_i and the P_i may be taken to be homogeneous.

c. Let $R = k[x, y]$, and let Γ be the abelian group $\mathbf{Z}/(2)$ with elements written 0 and 1. We give R a grading by Γ, letting R_0 be the set of polynomials whose terms all have even degree in y, and R_1 the set of all polynomials whose terms have odd degree in y. The element $x^2 - y^2$ is homogeneous of degree 0. Let $M = R/(x^2 - y^2)$. Show that M is also graded by Γ. Show that the prime ideal $P = (x - y)$ is the annihilator of an element of M, but that P is not homogeneous. (By part b, this shows that $\mathbf{Z}/(2)$ cannot be ordered in such a way that the action of $\mathbf{Z}/(2)$ on itself is compatible with the order. Prove this directly.) Show that $0 \subset M$ does not have a primary decomposition by homogeneous submodules of M.

Primary Decomposition of Monomial Ideals

Computing the primary decomposition of the ideal generated by an arbitrary set of polynomials is quite difficult. See for example Eisenbud, Huneke and Vasconcelos [1992] for algorithms and references. But for monomial ideals the job is relatively easy. See Heinzer, Ratliff, and Shah [in press] and Sturmfels, Trung, and Vogel [in press] for further information on monomial primary decomposition. See Eisenbud and Sturmfels [in press] for the case of binomial ideals.

Let k be a field (or any domain). A **monomial ideal** is an ideal $I \subset k[x_0, \ldots, x_r]$ generated by monomials in the variables $x_0, \ldots, x_r$.

Exercise 3.6:* Which monomial ideals are prime? Irreducible? Radical? Primary?

Exercise 3.7:* Find an algorithm for computing the radical of a monomial ideal.

Exercise 3.8:* Find an algorithm for computing an irreducible decomposition, and thus a primary decomposition, of a monomial ideal.

Exercise 3.9:* Products of linear primes

a. Let $I = (x_0) \cdot (x_0, x_1) \cdot \cdots \cdot (x_0, \ldots, x_r)$. Show that the associated primes of I are $(x_0), (x_0, x_1), \ldots, (x_0, \ldots, x_r)$.

b. More generally, for any subset $J \subset \{0, \ldots, r\}$, let $P(I)$ be the prime ideal generated by $\{x_i | i \in I\}$. Let $I_1, \ldots, I_t$ be subsets of $\{0, \ldots, r\}$, and set $I = \Pi_j P(I_j)$. Let Γ be the "incidence graph," whose vertices are the sets I_j, with an edge joining I_i and I_j iff $I_i \cap I_j \neq \emptyset$. Show that the associated primes of I are precisely those primes that may be expressed as $P(I_{j_1} \cup \cdots \cup I_{j_s})$ where $I_{j_1}, \ldots, I_{j_s}$ forms a connected subgraph of Γ. (It may not be easiest to use the general algorithm above.)

The Question of Uniqueness

Exercise 3.10:

a.* Let $R = k[a, b]/I$ where $I = (a) \cap (a, b)^2 = (a^2, ab)$. Show that (b^n) is (a, b)-primary in R, and that

$$0 = (a) \cap (b^n)$$

is a minimal primary decomposition of 0 in R for any $n \geq 1$.

b. Show that $(a + \lambda b^n)$ is also (a, b) primary for any nonzero $\lambda \in k$, and that

$$0 = (a) \cap (a + \lambda b^n).$$

Show that each $(a + \lambda b^n)$ is maximal among those ideals $J \subset R$ with

$$0 = (a) \cap J;$$

thus the length of the rings R/J, for J a "maximal (a, b)-primary component of 0," is actually unbounded.

c. It may be objected that example b is unnatural in the sense that it gives an inhomogeneous primary decomposition of a homogeneous ideal. However, it can be "homogenized" as follows: Let $S = R[c]$. Show that $0 = (a) \cap (ac^{n-1} + \lambda b^n)$ are primary decompositions of 0 in S, and that $(ac^{n-1} + \lambda b^n)$ is maximal among homogeneous ideals that can be used as primary components.

d.* (Huneke): For maximal associated primes in the homogeneous case there is a small positive result: Let $I \subset k[x_1, \ldots, x_r]$ be a homogeneous ideal and suppose that $R = k[x_1, \ldots, x_r]/I$ has the maximal ideal $(x_1, \ldots, x_r)$ as an associated prime. Show that there exists a number B such that if

$$I = J_1 \cap J_2 \cap \cdots$$

is a primary decomposition of I by homogeneous ideals, and J_1 is maximal among the homogeneous ideals that can appear as an $(x_1, \ldots, x_r)$-primary component, then the length of the ring R/J_1 is bounded above by B.

Exercise 3.11 (Uniqueness of maximal monomial primary decomposition):* (Bayer, Galligo, Stillman): Show that if $I \subset k[x_1, \ldots, x_r]$ is a monomial ideal, then there is a unique minimal primary decomposition $I = \cap I_j$ of I for which each I_j is a monomial ideal, primary to an ideal P_j generated by a subset of the variables, and I_j is maximal among the possible monomial P_j-primary components.

Exercise 3.12: Let M be a finitely generated module over the Noetherian ring R. Given any multiplicatively closed set $U \subset R$, show that the intersection of the primary components of 0 in M corresponding to those primes of Ass M not meeting U is the kernel of the localization map $M \to M[U^{-1}]$, and is thus independent of the primary decomposition chosen. Show that any such kernel may be written as $H^0_I(M)$ for some ideal $I \subset R$.

Determinantal Ideals

Exercise 3.13:

a. Let $M_r = \mathbf{A}^{r^2}$ be the affine space of $r \times r$ matrices over an algebraically closed field k. Show that if a polynomial f vanishes on all the matrices of rank s in M_r, then it must vanish on all matrices of rank $s - 1$.

b. Use part a and the idea of the proof given in the text for the case of 3×3 matrices of rank 1 to show that the set of $r \times r$ matrices of rank $\leq s$ is irreducible. (In fact, the ideal of $(s+1) \times (s+1)$ minors of the generic $r \times r$ matrix is prime —but this is somewhat harder to prove; see for example Bruns and Vetter [1988].)

c. Now show that if P is the radical of the ideal of $(s+1) \times (s+1)$ minors of the generic $r \times r$ matrix, then the $(s+2) \times (s+2)$ minors are in the symbolic square of P.

Total Quotients

Exercise 3.14: Use the finiteness of the set of associated primes of a Noetherian ring R to show that the total quotient ring $K(R)$ has only finitely many maximal ideals —they are the localizations of the maximal associated primes.

Exercise 3.15: The construction of the ring of total quotients $K(R)$ of a ring R (obtained from R by inverting all the nonzerodivisors of R) commutes with localization if the ring is reduced, but not in the general case. The problem has to do with embedded primes:

a.* If R is a reduced ring, show that for any multiplicatively closed set $U \subset R$ we have $K(R[U^{-1}]) = K(R)[U^{-1}]$.

b. If R is any ring and U is any multiplicatively closed subset, show that $K(R[U^{-1}]) = K(K(R)[U^{-1}])$ is a localization of $K(R)[U^{-1}]$.

c. Let k be a field, let $R = k[x,y,z]/(x^2, xy, xz)$, and let $P = (x,y)$. Show that

$$\begin{aligned} K(R) &= R_{(x,y,z)}; \\ R_P &= k[y,z]_{(y)}; \\ K(R_P) &= k(y,z); \end{aligned}$$

and thus $K(R_P) \neq R_P \otimes K(R) = R_P$.

Exercise 3.16: Give an example of an extension of finitely generated abelian groups for which the second inclusion of Lemma 3.6b is proper.

Prime Avoidance

Exercise 3.17: Here are two examples that show how the prime avoidance Lemma 3.3 can**not** be improved.

a. Show that if $k = \mathbf{Z}/(2)$ then the ideal $(x,y) \subset k[x,y]/(x,y)^2$ is the union of 3 properly smaller ideals.

b. Let k be any field. In the ring $k[x,y]/(xy,y^2)$, consider the ideals $I_1 = (x), I_2 = (y)$, and $J = (x^2,y)$. Show that the homogeneous elements of J are contained in $I_1 \cup I_2$, but that $J \not\subset I_1$ and $J \not\subset I_2$. Note that one of the I_j is prime.

Exercise 3.18: Prime avoidance usually fails for infinite sets of primes, but not always.

a. Show that in $k[x,y]$ the ideal (x,y) is contained in an infinite union of primes P_i such that no P_i contains (x,y).

b.* Suppose that $R = k[\{x_j\}_{j\in A}]$ is a polynomial ring with infinitely many variables indexed by a set A. Let $\{A_i\}_{i\in B}$ be a (possibly) infinite collection of mutually disjoint subsets of A, and for each $i \in B$ let P_i be the prime ideal generated by $\{x_j\}_{j\in A_i}$. Show that any ideal contained in the union of the P_i is contained in one of them. Conclude that if U is the multiplicative set $U = R - \cup_{i\in B} P_i$, then the maximal ideals of $S = R[U^{-1}]$ are precisely the ideals SP_i. See Exercise 9.6 for more about this example.

Exercise 3.19 (Refinements of prime avoidance):* Prove the following useful variants of prime avoidance:

a. Suppose R is a ring containing a field k, and let $I_1, \dots, I_n$ be ideals of R. If $(f_1,\dots,f_n) \not\subset I_i$ for $i = 1,\dots,s$, then there is a nonzero homogeneous polynomial $g(t_1,\dots,t_n) \in k[t_1,\dots,t_n]$ with the property that

(*) $\sum a_i f_i \notin \bigcup_j I_j$ for all $(a_1,\dots,a_n) \in k^n$ such that $g(a_1,\dots,a_n) \neq 0$.

In particular, If k is infinite, then there is an element $(a_2,\dots,a_n) \in k^{n-1}$ so that $f_1 + \sum_{i=2}^n a_i f_i \notin \cup_j I_j$.

b. Suppose R is a ring, and let $I_1,\dots,I_n$ be prime ideals of R. If $f \in R$ and J is an ideal of R such that $f + J \not\subset I_i$ for $i = 1,\dots,n$, then there is an element $g \in J$ with the property that

$$f + g \notin \bigcup_j I_j.$$

In particular, if $(f_1,\dots,f_s) \not\subset I_i$ for $i = 1,\dots,n$, then there is an element $(a_2,\dots,a_n) \in R^{n-1}$ so that $f_1 + \sum_{i=2}^n a_i f_i \notin \cup_j I_j$.

Exercise 3.20: Let M be a finitely generated module M over a Noetherian ring R. Proposition 3.4 immediately implies that the set of elements of R that are zerodivisors on M is a union of primes. Here is a method, due to Kaplansky, for showing directly that this set is a finite union of primes: Consider

$$\mathfrak{p} = \{(P,m) | P \text{ is a maximal annihilator ideal and } P = \operatorname{ann} m\}.$$

Let $M' \subset M$ be the submodule generated by all the m that occur as second members of pairs in $\mathfrak{p}$. Let $m_1, \ldots, m_n$ be a finite set of these m_i that generate M', and let $P_1, \ldots, P_n$ be the corresponding primes. Show that the set of zerodivisors on M is $P_1 \cup \cdots \cup P_n$.

4

Integral Dependence and the Nullstellensatz

The problem of solving equations and saying something about the solutions is a fundamental motivation and goal of commutative algebra. In pursuing this goal, it is often important to adjoin a solution of a polynomial equation in one variable: Given a ring R and a polynomial $p(x) \in R[x]$, the ring $R[x]/(p)$ may be thought of as the result of adjoining a root of p to R as freely as possible; the root adjoined is of course the image of x.

The study of localization and its cousin, primary decomposition, which has occupied us for the last two chapters, may be regarded as the study of the case where p is a linear polynomial with unit constant coefficient, which we might as well write as $p(x) = ax - 1$. In this chapter we shall take up another central case, in which p is a **monic** polynomial, that is, a polynomial $p(x) = x^n + r_1x^{n-1} + \cdots + r_n$ whose leading coefficient is 1. This case may be distinguished by the following fundamental remark, whose proof we give later as an application of the Cayley-Hamilton theorem. (See Exercises 3.4b and 6.5 for related results.)

Proposition 4.1. *Let R be a ring and let $J \subset R[x]$ be an ideal in the polynomial ring in one variable over R. Let $S := R[x]/J$, and let s be the image of x in S.*

a. S is generated by $\leq n$ elements as an R-module iff J contains a monic polynomial of degree $\leq n$. In this case S is generated by $1, s, \ldots, s^{n-1}$. In particular, S is a finitely generated R-module iff J contains a monic polynomial.

b. *S is a finitely generated free R-module iff J can be generated by a monic polynomial. In this case S has a basis of the form $1, s, \ldots, s^{n-1}$.*

If S is an R-algebra, and $p(x)$ is a polynomial with coefficients in R, then we say that an element $s \in S$ **satisfies p** if $p(s) = 0$. The element s is called **integral over R** if it satisfies a monic polynomial with coefficients in R. The equation $p(s) = 0$ is then called an **equation of integral dependence** or an **integral equation** for s over R. If every element of S is integral over R, we say that S itself is integral over R. The following result is the second key fact that makes this theory interesting.

Theorem 4.2. *Let R be a ring and let S be an R-algebra. The set of all elements of S integral over R is a subalgebra of S. In particular, if S is generated by elements integral over R, then S is integral over R.*

In particular, Theorem 4.2 shows that the algebra obtained by adjoining the solutions to any set of integral equations is integral. The proof of Theorem 4.2 will also be given later as an application of the Cayley-Hamilton theorem.

Given an R-algebra S, the ring of all elements of S integral over R is called the **integral closure**, or **normalization of R in S**. The most important examples occur when R is an integral domain and S is its quotient field. In this case the subalgebra of elements of S integral over R is simply called the **normalization of R**. A domain equal to its own normalization is called a **normal domain**.

Generalizing the normalization of a domain in its quotient field, an R-algebra S containing a copy of R as $R \cdot 1$ is called an **integral extension** of R if every element of S satisfies a monic polynomial with coefficients in R.

Integral extensions and normalization appear naturally in many contexts. For example:

- Geometrically, integral extensions of affine rings correspond to the maps of affine algebraic sets that are finite and **proper**. (Over the complex numbers, this means that the preimage of every set that is compact in the classical topology is again compact; in general it has a formulation that we shall explain in Chapter 14). If $\varphi : X \to Y$ is a morphism of algebra varieties, then the set of connected components of fibers of φ form an algebraic set Y' mapping to Y through which φ factors (the "Stein factorization of a morphism"; see Grothendieck [1961]).

- By an important finiteness result of Emmy Noether (Corollary 13.13), normalization is an operation that takes affine rings to affine rings. Since it also commutes with localization, it extends easily to an operation on any algebraic variety. It has the effect of "improving" certain

irregularities in a variety, and it is an important step toward resolution of singularities.

- There is a criterion, due to Serre, to test when a ring is equal to its normalization in terms of certain geometric and homological properties of the ring. A natural extension gives one of the most important tests for the primeness of an ideal. See Theorem 18.15 and the discussion following it.

- To return to our roots (pun intended), both the rings $\mathbf{Z}[x]/(x^2+1) \cong \mathbf{Z}[i] \subset \mathbf{Q}[i]$ and $\mathbf{Z}[x]/(x^2+4) \cong \mathbf{Z}[2i] \subset \mathbf{Q}[i]$ are interesting because they reflect the arithmetic in $\mathbf{Z}$. But the ring $\mathbf{Z}[i]$ is "nicer" than the ring $\mathbf{Z}[2i]$. For example, the first ring has unique prime factorization whereas the second does not, since $(2i)(2i) = -(2)(2)$. It will turn out that the ring $\mathbf{Z}[i]$ is the normalization of $\mathbf{Z}[2i]$.

As a slightly less obvious example, consider the ring $R = \mathbf{Z}[\sqrt{5}] = \mathbf{Z}[1+\sqrt{5}] \subset \mathbf{Q}$ and the larger ring $S = \mathbf{Z}[1/2 + 1/2\sqrt{5}]$. In the first ring the equation $(1+\sqrt{5})(1-\sqrt{5}) = -4 = -(2)(2)$ suggests that R does not have unique factorization. An easy check shows that this is true (one must show that $1+\sqrt{5}$, $1-\sqrt{5}$, and 2 cannot be factored further, and that they do not differ by units of R). However, in S we see that $1+\sqrt{5} = 2(1/2 + 1/2\sqrt{5})$ and $1-\sqrt{5} = 2(1/2 - 1/2\sqrt{5}) = 2(1/2 + 1/2\sqrt{5} - \sqrt{5})$. Also, $(1/2+1/2\sqrt{5})(1/2-1/2\sqrt{5}) = -1$, so both $1/2+1/2\sqrt{5}$ and $1/2-1/2\sqrt{5}$ are units. Thus the two factorizations are essentially the same in S; in fact, one can show that S has unique factorization into primes. Again, the reason that S is "better" than R is that S is the normalization of R. Although even normal rings of algebraic integers may fail to have unique factorization, we shall see that they always have it locally, whereas non-normal ones do not.

In general, if K is a number field (that is, a finite extension field of $\mathbf{Q}$), then the set of elements of K that satisfy monic equations with coefficients in $\mathbf{Z}$ is called the ring of **algebraic integers** in K. As we remarked in Chapter 1, these are the rings whose study started commutative algebra.

4.1 The Cayley-Hamilton Theorem and Nakayama's Lemma

The classical Cayley-Hamilton theorem says that a linear transformation on a finite-dimensional vector space satisfies its characteristic polynomial. Hamilton proved this for linear transformations of $\mathbf{R}^3$ in 1853, and Cayley announced the general case in 1858, though he too seems only to have checked the cases up to 3×3 matrices. For purposes later in this chapter, we shall need a more general version.

Theorem 4.3 (Cayley-Hamilton). *Let R be a ring, $I \subset R$ an ideal, and M an R-module that can be generated by n elements. Let φ be an endomorphism of M. If*

$$\varphi(M) \subset IM,$$

then there is a monic polynomial

$$p(x) = x^n + p_1 x^{n-1} + \cdots + p_n$$

with $p_j \in I^j$ for each j, such that $p(\varphi) = 0$ as an endomorphism of M.

Despite the generality, the proof is virtually the same as for the classical case.

Proof. Let $m_1, \ldots, m_n$ be a finite set of generators of M. We may write each $\varphi(m_i)$ in terms of the m_j, using coefficients in I:

$$\varphi(m_i) = \Sigma a_{ij} m_j, \qquad \text{with } a_{ij} \in I.$$

We regard M as a module over the polynomial ring $R[x]$ by letting x act as φ. Let A be the $n \times n$ matrix with entries a_{ij}, and let $\mathbf{1}$ be the $n \times n$ identity matrix. If we write m for the column vector whose entries are the m_j, then the equations above say that

$$(x\mathbf{1} - A) \cdot m = 0.$$

Multiplying the left-hand side by the matrix of cofactors of $x\mathbf{1} - A$, we get

$$[\det(x\mathbf{1} - A)]\mathbf{1} \cdot m = 0,$$

that is, $\det(x\mathbf{1} - A)m_i = 0$ for all i; thus

$$[\det(x\mathbf{1} - A)]M = 0.$$

It follows that the polynomial $p(x) = \det(x\mathbf{1} - A)$ has the desired property $p(\varphi) = 0$. It is easy to see directly that the jth coefficient p_j is in the jth power of I; from a high-brow point of view, this is because p_j is the trace of the jth exterior power of A, the sum of certain $j \times j$ minors of A. □

Before returning to the application to integral elements, we give a useful and surprising application in another direction.

Corollary 4.4. *Let R be a ring, and let M be a finitely generated R-module.*

a. *If $\alpha : M \to M$ is an epimorphism of R-modules, then α is an isomorphism.*

b. *If $M \cong R^n$, then any set of n elements of M that generate M forms a free basis; in particular, the rank n of M is well defined.*

Proof.

a. We may regard M as a module over $R[t]$, letting t act by $tm = \alpha(m)$ for $m \in M$. If we set $I = (x) \subset R[t]$, then since α is an epimorphism, $IM = M$. Thus we may apply the Cayley-Hamilton theorem with φ the identity endomorphism of M. It follows that there is a polynomial $q(t)$ such that $(1 - q(t)t)M = 0$, or equivalently $1 - q(\alpha)\alpha = 0$. From this we see that $q(\alpha)$ is the inverse to α, and α is an isomorphism.

b. A set of n generators of M corresponds to a surjection $\beta : R^n \to M$ sending the basis elements of R^n to the given generators of M. Since M is free of rank n, we may choose an isomorphism $\gamma : M \to R^n$. The map $\beta\gamma : M \to M$ is a surjection, and thus an isomorphism. It follows that $\beta = (\beta\gamma)\gamma^{-1}$ is an isomorphism, so the given generators for M are a free basis.

   To prove that the rank of a finitely generated free module is well defined, suppose that $R^m \cong R^n$. If $m \neq n$, suppose that $m < n$. We can extend a basis of length m by adjoining some elements equal to zero, to obtain a set of n generators that do not form a free basis, contradicting the first statement of part b. Thus $m = n$, and we see that the rank is well defined. (One could prove this last statement directly: If p is a maximal ideal of R then $(R/P)\otimes_R R^m = (R/P)^m$ is a vector space of dimension m. By the same argument it has dimension n, so $m = n$). □

The criterion of Corollary 4.4a is often useful when one can "approximate" a homomorphism in some way; see Exercises 4.13 and 7.5 for examples. Corollary 4.4b is, of course, fundamental. See Exercise 4.10 for a different proof. This statement is not so trivial as it might seem: The rank is not, in general, a well-defined invariant of a free module over a noncommutative ring (although it is well defined in the Noetherian case). For example if an abelian group A satisfies $A \cong A \oplus A$, as does for example any infinite dimensional vector space, and $\Gamma := \operatorname{Hom}(A, A)$ is its ring of endomorphisms, then $\Gamma = \operatorname{Hom}(A, A) = \operatorname{Hom}(A, A \oplus A) = \Gamma \oplus \Gamma$ as right-Γ-modules. The trick used here is sometimes called the "Eilenberg Swindle".

Next we use these results to prove Proposition 4.1.

Proof of Proposition 4.1.

a. The powers of x generate $R[x]$ as an R-module, so their images, the powers of s, generate S. Suppose that J contains a monic polynomial p of degree n. Any power s^d of s with $d \geq n$ may be written in terms of smaller powers by means of the equation $0 = s^{d-n}p(s) = s^d + r_1 s^{d-1} + \cdots$, so the first n powers of s generate S.

   Conversely, suppose that S can be generated as an R-module by n elements. We regard multiplication by s as an endomorphism of the

R-module S. Taking $I = R$, the Cayley-Hamilton theorem shows that s satisfies a monic polynomial $p(x)$ of degree n. Since $p(s) = 0$, the polynomial $p(x)$ is in J.

b. Suppose that J is generated by a monic polynomial p of degree n. We know from part a that the first n powers of s generate S. To show they are linearly independent, suppose $\sum_0^{n-1} a_i s^i = 0$ for some elements $a_i \in R$. It follows that the polynomial $q(x) = \sum_0^{n-1} a_i x^i$ is in $J = (p)$. Since p is monic, any nonzero multiple of p has degree equal to n or greater than n, and we see that $q = 0$. This shows that S is a free R-module having the first n powers of s as free basis.

Conversely, suppose S is a free R-module, and let n be its rank. As an R-module, S can generated by n elements, so by part a there is a monic polynomial p of degree n in J. It follows by part a that S is generated as an R-module by $1, \ldots, s^{n-1}$. Since S is free of rank n, Corollary 4.4b shows that these powers form a basis of S as an R-module.

We claim that p generates J. If $f \in J$ is any polynomial, let q be the remainder of f on division by p, so that $q \in J$ and degree $q < n$. As above, the polynomial q can be interpreted as a linear relation among some of the first n powers of s. Since these form a free basis of S, it follows that $q = 0$, so f is divisible by p as required. □

In general we shall say that an R-algebra S that is finitely generated as an R-module is **finite over R**. This is stronger than being integral. The following result extends the connection given in Proposition 4.1 to rings generated by more than one element.

Corollary 4.5. *An R-algebra S is finite over R iff S is generated as an R-algebra by finitely many integral elements.*

Proof. First suppose that S is finite over R. If $s \in S$, then multiplication by s is an endomorphism of S, and the Cayley-Hamilton theorem shows that s satisfies an integral equation.

For the converse, suppose S is generated by t elements, and let S' be the subalgebra of S generated by $t - 1$ of the generators. We may assume by induction that S' is finite over R. Suppose S' is generated as an R-module by a finite set of elements $\{s_i\}$. The last generator, s, is integral over R and thus also integral over S'; by Proposition 4.1 there is a finite set of generators of S as an S'-module, say $\{t_j\}$. It is easy to check that the set of products $\{s_i t_j\}$ generates S as an R-module. □

For the natural generality of the idea in the second part of the proof, see Exercise 4.1.

Here is the application of the Cayley-Hamilton theorem, Theorem 4.3, to integral elements.

Corollary 4.6. *If S is an R-algebra and $s \in S$ then s is integral over R iff there exists an S-module N and a finitely generated R-submodule $M \subset N$, not annihilated by any nonzero element of S, such that $sM \subset M$. In particular, s is integral iff $R[s]$ is a finitely generated R-module.*

Proof. Suppose first that s is integral over R. Take $N = S$. By Proposition 4.1, $M = R[s] \subset S$ is finitely generated as an R-module.

Conversely, we may regard multiplication by s as an endomorphism of M. Applying the Cayley-Hamilton theorem we see that there is a monic polynomial p having coefficients in R with $p(s)M = 0$. From our hypothesis it follows that $p(s) = 0$ as an element of S, and thus s is integral as required.

The last statement, which may also be regarded as a restatement of Proposition 4.1, follows because $1 \in R[s]$ is not annihilated by any nonzero element of S. □

It would be natural to prove Theorem 4.2 by starting with the equations satisfied by two integral elements and simply writing down the equations satisfied by their sum and product. In a sense, this is what we shall do. But in general the necessary polynomials are complicated. The Cayley-Hamilton theorem gives them implicitly.

Proof of Theorem 4.2. Let s, s' be elements of S that are integral over R. We must show that $s + s'$ and ss' are integral over R. Suppose that $M = R[s]$, and $M' = R[s'] \subset S$. By Proposition 4.1 both M and M' are finitely generated modules. We define MM' to be the module spanned by all the pairwise products of elements of M and M'. Since it would be enough to take pairwise products of generators of M and M', the module MM' is also a finitely generated module. We have

$$\begin{aligned} ss'MM' &= sMs'M' \subset MM' \\ (s+s')MM' &\subset sMM' + Ms'M' \subset MM' + MM' = MM', \end{aligned}$$

so both ss' and $s + s'$ are integral by Corollary 4.6. This shows that the integral elements form a subring. □

If R is Noetherian, then in fact Theorem 4.2 follows directly from Corollary 4.5. Replacing S by the algebra generated by the elements integral over R, it suffices to show that every element of S is integral over R. If $s \in S$, then s is in a subring S' generated by just finitely many integral elements, which is finite by Proposition 4.5. Since R is Noetherian the subalgebra $R[s] \subset S'$ is finite over R, so s is integral by the other implication of Corollary 4.5.

We now give two further consequences of the Cayley-Hamilton theorem that will be of great importance in the next chapters.

Corollary 4.7. *If M is a finitely generated R-module and I is an ideal of R such that $IM = M$, then there is an element $r \in I$ that acts as the identity on M; that is, such that $(1 - r)M = 0$.*

Proof. Take φ to be the identity in Lemma 4.3; the resulting equation $p(1)M = 0$ becomes

$$(1 + p_1 + \cdots + p_n)M = 0,$$

with p_j in $I^j \subset I$, so we may take $r = -(p_1 + \cdots + p_n)$. □

The next result, called Nakayama's lemma (see the history in Nagata [1962] p. 212–213), is an extraordinarily useful tool in the theory of local rings. To state it in maximal generality we use the following definition.

Definition. *The **Jacobson radical** of a ring R is the intersection of all the maximal ideals of R.*

Corollary 4.8 (Nakayama's Lemma). *Let I be an ideal contained in the Jacobson radical of a ring R, and let M be a finitely generated R-module.*

a. If $IM = M$, then $M = 0$.

b. If $m_1, \ldots, m_n \in M$ have images in M/IM that generate it as an R-module, then $m_1, \ldots, m_n$ generate M as an R-module.

Proof.

a. We apply Corollary 4.7 to get $r \in I$ such that $(1 - r)M = 0$. Since r is in every maximal ideal, $1 - r$ is in no maximal ideal; that is, $1 - r$ is a unit. It follows that $M = 0$.

b. Let $N = M/(\sum_i Rm_i)$. We have $N/IN = M/(IM + (\sum_i Rm_i)) = M/M = 0$, so $IN = N$. We now apply part a to get $N = 0$, that is, $M = (\sum_i Rm_i)$. □

Warning: It is tempting, but in general wrong, to use Nakayama's lemma to prove that a module M is finitely generated by exhibiting finitely many generators for M/IM. But in some favorable cases this argument is correct; see Exercises 4.6 and 7.2.

Nakayama's Lemma says in particular that if (R, P) is a local ring and M is a finitely generated R-module such that $M/PM = 0$, then $M = 0$. Since $M/PM = R/P \otimes M$, the following result extends this remark.

Corollary 4.9. *If M and N are finitely generated modules over a ring R, and $M \otimes_R N = 0$, then $\operatorname{ann} M + \operatorname{ann} N = R$. In particular, if R is local, then either M or N is 0.*

Proof. It suffices to prove the local case, since if $\operatorname{ann} M + \operatorname{ann} N \neq R$, we could localize at a prime ideal containing both $\operatorname{ann} M$ and $\operatorname{ann} N$, and apply the local result to get a contradiction. Assuming that (R, P) is local, and $M \neq 0$, Nakayama's lemma implies that $M/PM \neq 0$. Since this is an R/P vector space, it projects onto R/P and so there is a surjection from M itself onto R/P. Thus $0 = M \otimes N$ surjects onto $R/P \otimes N = N/PN$. By Nakayama's lemma, $N = 0$. □

4.2 Normal Domains and the Normalization Process

We have already hinted that there is a connection between normality and unique factorization. The following proposition gives the relation.

Proposition 4.10. *Let R be a ring. If R is factorial, then R is normal.*

Proof. Suppose that R is factorial, and that r/s with $r, s \in R$, is a fraction that is integral over R. We may assume that r and s are relatively prime, and we wish to show that $r/s \in R$. If the integral equation satisfied by r/s is

$$(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots = 0,$$

then multiplying by s^n gives

$$r^n + sa_{n-1}r^{n-1} + \cdots = 0.$$

Thus r^n is divisible by s, contradicting the relative primeness of r and s.□

Proposition 4.10 shows immediately that the domain $\mathbf{Z}$ is normal. If k is a field then $k[x_1, \ldots x_r]$ and $\mathbf{Z}[x_1, \ldots, x_r]$ are factorial, so these rings are normal too. In fact more is true: A ring R is normal iff the polynomial ring $R[x]$ is normal—see Exercise 4.18.

If $R \subset S$ are rings and $f(x) \in R[x]$ is a monic polynomial with a root in S, then by definition the root is integral over R. Having a root α is the same as having a linear factor $(x - \alpha)$. The following result shows that something similar is true for any factor, linear or not. Given that $\mathbf{Z}$ is normal, as we have just shown, it generalizes the statement that a monic polynomial with integer coefficients that is irreducible in $\mathbf{Z}[x]$ is also irreducible in $\mathbf{Q}[x]$. (This is usually proved from Gauss' lemma; see Exercise 3.4.)

Proposition 4.11. *Let $R \subset S$ be rings, and suppose that $f \in R[x]$ is a monic polynomial. If f factors in $S[x]$ as $f = gh$, with g and h monic, then the coefficients of g and h are integral over R.*

Proof. Adjoining a root α_1 of g to S and using long division in the ring $S[\alpha_1] = S[x]/(g)$, we see that g factors as $(x - \alpha_1)g_1$, where the degree of g_1 is one less than the degree of g. Repeating this process inductively, we may find an extension ring T of S and elements α_i and β_j of T such that $g = \Pi(x - \alpha_i)$, $h = \Pi(x - \beta_j)$ in $T[x]$. Since each α_i and β_j is a root of the monic polynomial f, the subring T' of T generated as an R-algebra by the α_i and β_j is integral over R. Since the coefficients of g and h are the elementary symmetric functions in the α_i and the β_j, respectively, they too are integral over R. □

If R is a domain and $f = gh \in R[x]$ is a factorization of a monic polynomial into nonmonic polynomials, then because f is monic the leading coefficients of g and h are units of S, inverse to one another. Multiplying each of g and h by the leading coefficient of the other produces a factorization to which Proposition 4.11 applies.

As a consequence we get a weak converse of Proposition 4.10, tightening the connection of normality and factoriality and generalizing another standard consequence of Gauss' lemma. It is useful with such results as Eisenstein's criterion, Exercise 18.11.

Corollary 4.12. *If R is a normal domain, then any monic irreducible polynomial in $R[x]$ is prime.*

Proof. Let f be a monic irreducible polynomial. Write Q for the quotient field of R. By Proposition 4.11, f remains irreducible in $Q[x]$. Since $Q[x]$ is factorial, $P = fQ[x]$ is prime. Since $R[x]/(f)$ is free over R, the map $R[x]/(f) \to Q \otimes_R R[x]/(f) = Q[x]/P$ is a monomorphism; thus (f) is prime in $R[x]$. □

Normalization commutes with localization.

Proposition 4.13. *Let $R \subset S$ be rings, and let U be a multiplicatively closed subset of R. If S' is the integral closure of R in S, then $S'[U^{-1}]$ is the integral closure of $R[U^{-1}]$ in $S[U^{-1}]$.*

Proof. An element of S integral over R is certainly integral over $R[U^{-1}]$, so $S'[U^{-1}]$ is integral over $R[U^{-1}]$. For the other inclusion we must show that if $s/u \in S[U^{-1}]$ is integral over $R[U^{-1}]$ then s times an element of U is integral over R. If

$$(s/u)^n + (r_1/u_1)(s/u)^{n-1} + \cdots = 0$$

is an equation of integrality for s/u, then we can clear denominators by multiplying by $(uu_1 \cdots u_n)^n$, to get a relation of integrality for $su_1 \cdots u_n$:

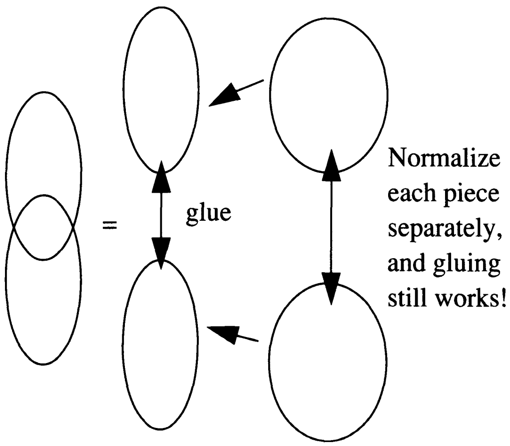
$$(su_1 \cdots u_n)^n + r_1(uu_2 \cdots u_n)(su_1 \cdots u_n)^{n-i} + \cdots = 0. \qquad \square$$

If R is a Noetherian domain then one might hope that the integral closure S of R (in its quotient field or perhaps some finite extension field) would be Noetherian. If S is finitely generated as an R-algebra, then this will true by Corollary 4.5. In general, an integral algebra over a ring R is a possibly infinite union of finite algebras. If this union is really infinite, then S might not be Noetherian; this can really happen, as an example due to Nagata [1962, Example 8, p. 211] shows. However, in the case of affine rings, all is well.

Theorem 4.14 (Emmy Noether). *If R is a finitely generated domain over a field or over the integers, and L is a finite extension field of the field of fractions of R, then the integral closure of R in L is a finitely generated R-module.*

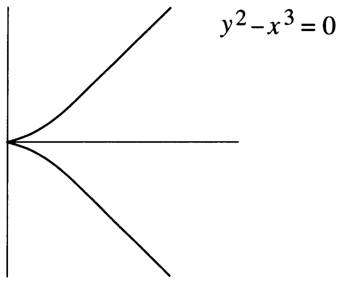
We shall give the proof of Theorem 4.14 in the case of affine rings as an application of the Noether normalization theorem in Corollary 13.13 (for the general case see the references given there). In a similar vein, the Krull-Akizuki theorem (Theorem 11.13) shows that the normalization of a "one-dimensional" Noetherian ring is again Noetherian. In general, rings R satisfying the conclusion of Theorem 4.14 were named **Japanese rings** by Grothendieck [1965] in honor of the contributions of the Japanese school.

Theorem 4.14, especially together with Proposition 4.13, is useful in geometry: Theorem 4.14 says essentially that the normalization of an affine variety is again an affine variety, and it is generally somewhat simpler. Proposition 4.13 shows that a general reduced, irreducible variety made by patching together affine pieces (for example, a projective variety) has a nice normalization: We may normalize each affine piece separately, and then glue together along the open sets (see the discussion in Chapter 2), which are the normalizations of the open sets along which the original pieces were glued.

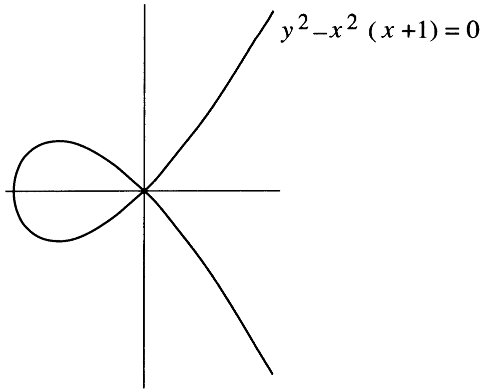


4.3 Normalization in the Analytic Case

There is a beautiful interpretation of integrality for rational functions on an affine variety over **C** that comes from complex analysis. We shall sketch a special case, referring the reader to Gunning and Rossi [1965] for the general case. For simplicity, we assume that the variety is regular except for one singular point (see Chapter 10 and Theorem 16.19 for the definition, or just look at the pictures below). In this case a rational function is integral over the ring of polynomial functions localized at the singular point iff it stays bounded in a neighborhood of the singular point in the classical topology. For example, consider the coordinate rings $\mathbf{C}[x,y]/(y^2-x^3)$ and $\mathbf{C}[x,y]/(y^2-x^2(x+1))$ of the plane curves



and



respectively. In each case the singular point is the origin (0,0). In each case the function y/x, though not regular in the sense that it is in the coordinate ring (and not bounded in a neighborhood of the singular point in the plane), does stay bounded along the curves. (For the real points of the curves one sees this plainly from the pictures: It just means that the distance from a point on the curve to the x-axis is never too much greater than the distance of the point to the y-axis. For the complex points a little algebra is necessary, as usual.)

In each of these cases, it is easy to see the integral equation satisfied by this bounded function: In the first case it is $(y/x)^2 - x = 0$, in the second $(y/x)^2 - (x-1) = 0$. The "reason" why y/x is not in the coordinate ring is that while all polynomial functions are restrictions of polynomial functions on the plane, the function y/x does not even extend to a continuous (in the classical sense) function on the plane. In the second case, y/x is not even continuous on the curve: Along one "branch" of the curve, y/x has limit 1 at $(0,0)$, whereas along the other branch it has limit -1. Thus it "separates" the two branches; see Exercise 4.24.

4.4 Primes in an Integral Extension

Suppose that R and S are affine k-algebras corresponding to varieties X and Y, respectively. A homomorphism $R \to S$ corresponds to a morphism $Y \to X$. The homomorphism $R \to S$ is an inclusion iff no polynomial function on X pulls back to 0 on Y—that is, iff the image of $Y \to X$ is not contained in any proper closed subset of X or, in fancier language, iff the image is dense in the Zariski topology.

What does it mean for S to be integral over R? The full answer, which we shall not completely explain here, is that the map $Y \to X$ is **proper** with finite fibers. If the ground field is $\mathbf{C}$, then this means that the preimage in Y of a compact subset of X (in the classical topology) is a compact subset of Y; over a general ground field, properness is a good replacement for this sort of relative compactness. We shall meet properness again when we come to elimination theory in Chapter 14; the interested reader can find a technical account of the general notion in Hartshome [1977, Chapter 2]. For now we shall prove three facts, formulated by Cohen and Seidenberg [1946], that reflect part of this geometric constellation. Their main use is to show that if $R \subset S$ is an integral extension, then chains of prime ideals in R and in S are closely related; such information will be necessary when we come to dimension theory. We shall also use Corollary 4.17 in the proof of the Nullstellensatz at the end of this chapter. All three results are essentially corollaries of Nakayama's lemma.

Proposition 4.15 (Lying Over and Going Up). *Suppose that $R \subset S$ is an integral extension of rings. Given a prime P of R, there exists a prime Q of S with $R \cap Q = P$, and in fact Q may be chosen to contain any given ideal Q_1 that satisfies the (obviously necessary) condition $R \cap Q_1 \subset P$.*

The first statement of the proposition is called *Lying Over* because it asserts the existence of a prime of S "lying over" a given prime of R. The second statement is called *Going Up* because it constructs a prime Q "up" from Q_1 (see Figure 4.1). There is also a somewhat deeper *going down* result that holds under stronger hypotheses (Theorem 13.9).

Proof. Factoring out Q_1 and $R \cap Q_1$, we may suppose that $Q_1 = 0$, and we need only prove the existence of a prime Q of S with $R \cap Q = P$. Let U be the multiplicatively closed set $R - P$. Replacing R by $R_P = R[U^{-1}]$ and S by $S[U^{-1}]$, we may assume that R is local with maximal ideal P.

With these hypotheses, any maximal ideal of S containing PS has preimage containing P, and therefore equal to P; so we need only prove that $PS \neq S$. But if $PS = S$ then $1 \in S$ can be expressed as an S-linear combination of finitely many elements of P. If we let S' be the subalgebra of S generated by these elements, then $1 \in PS'$ so $PS' = S'$. Since S and thus S' are integral over R, Corollary 4.5 shows that S' is a finitely generated R-module. By Nakayama's lemma, $S' = 0$, a contradiction. □

Even if we only assume that the quotient field of S is algebraic over that of R, the following lemma shows that there is some relation between the ideal theory of S and that of R. Recall from Chapter 2 that if R is a domain, then $K(R)$ denotes the quotient field of R.

Lemma 4.16. *Let $R \subset S$ be domains. If $K(S)$ is algebraic over $K(R)$ then any nonzero ideal of S intersects R nontrivially.*

Proof. It suffices to treat a principal ideal bS. Now b satisfies an equation of the form $a_n b^n + \cdots + a_1 b + a_0 = 0$, with each $a_i \in K(R)$. Multiplying by a common denominator of the a_i in R, and dividing by a power of b if necessary, we may suppose all $a_i \in R$ and $a_0 \neq 0$. Clearly, $a_0 \in bS$. □

With a hypothesis as in Lemma 4.16, suppose that R is a field. By the lemma, any nonzero ideal of S contains a nonzero element of R, and this is a unit; thus S is a field. If we assume that S is integral over R, then the converse is true as well.

Corollary 4.17. *If $R \subset S$ is an integral extension of domains, then S is a field iff R is a field. Equivalently, if S is an integral R-algebra and P is a prime of S, then P is a maximal ideal of S iff $P \cap R$ is a maximal ideal of R.*

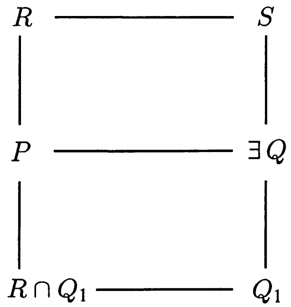


FIGURE 4.1.

Proof. We have already seen that if R is a field then S is too. For the converse, if $\mathfrak{m}$ is a maximal ideal of R, then by Lying Over there is a prime Q of S intersecting R in $\mathfrak{m}$. If S is a field, then $Q = 0$, so $\mathfrak{m} = R \cap Q = 0$, and R is a field. The second statement of the corollary may be reduced to the first by factoring out P and $P \cap R$. □

For a direct proof of Corollary 4.17, see Exercise 4.3.

The hypothesis of Lemma 4.16 may be spoiled if we pass to factor rings of R and S, since the equations that make the elements of S algebraic over R may become trivial. However, integral equations cannot become trivial in such a factor ring, and this makes the lemma particularly potent with a hypothesis of integrality. Here is a typical application; for another, see Theorem 11.13.

Corollary 4.18 (Incomparability). *Suppose $R \subset S$ is an integral extension of rings. Two distinct primes of S having the same intersection with R are incomparable.*

Proof. If $Q \subset Q_1 \subset S$ are primes, with $R \cap Q = R \cap Q_1 = P \subset R$, then factoring out P in R and Q in S reduces to a situation where S is a domain, $Q = 0$, and $Q_1 \cap R = 0$. Since integral equations persist modulo P, S is still integral over R, and thus $K(S)$ is algebraic over $K(R)$. Lemma 4.16 shows that $Q_1 = 0 = Q$, as required. □

Sometimes more is true: We shall show in Proposition 3.10 that if R is normal and $K(S)/K(R)$ is Galois, then any two primes with the same preimage in R are conjugate under an automorphism of S.

4.5 The Nullstellensatz

The original Nullstellensatz, explained in the Introduction, deals with polynomials in n variables over a field. The version below is much more general: It deals with a property that fields (trivially) possess and that is preserved by every polynomial extension. It does not even need a Noetherian hypothesis. At the end of this chapter we shall show explicitly how to derive the version of the Nullstellensatz given in Chapter 1 from the version given here. The exercises contain three further proofs; a fifth will be given in Chapter 13.

We say that a ring R is a **Jacobson ring** if every prime ideal of R is the intersection of maximal ideals[1]. It is obvious that any field is a Jacobson ring. The version of the Nullstellensatz that we shall now prove, shows that

[1] This name, bestowed by Krull, honors Nathan Jacobson's studies of the intersection of the maximal ideals of a ring, which is now called the Jacobson radical. The name "Hilbert ring" also appears in the literature.

any finitely generated algebra over a field is also a Jacobson ring. Recall that if S is an R-algebra by some homomorphism $\alpha : R \to S$, and $I \subset S$ is an ideal, then for simplicity we write $I \cap R$ for $\alpha^{-1}(I)$, just as if α were an inclusion.

Theorem 4.19 (Nullstellensatz—General form). *Let R be a Jacobson ring. If S is a finitely generated R-algebra, then S is a Jacobson ring. Further, if $\mathfrak{n} \subset S$ is a maximal ideal, then $\mathfrak{m} := \mathfrak{n} \cap R$ is a maximal ideal of R, and $S/\mathfrak{n}$ is a finite extension field of $R/\mathfrak{m}$.*

The conclusion of the second statement can easily fail if R is not Jacobson (the conclusion of the first statement fails trivially!). For example, let $R = k[t]_{(t)}$. The unique maximal ideal of R is (t), so the prime ideal 0 is not an intersection of maximal ideals of R, and R is not Jacobson. If $\mathfrak{n} = (xt-1) \subset S := R[x]$, then $S/\mathfrak{n} \cong k(t)$, so $\mathfrak{n}$ is a maximal ideal of S, but $\mathfrak{n} \cap R = 0$.

The proof will make use of a reformulation of the Jacobson property, variants of which go under the name "Rabinowitch's trick" (Rabinowitch [1929]).

Lemma 4.20. *Let R be a ring. The following are equivalent:*

a. R is Jacobson.

b. If P is a prime of R and if $S := R/P$ contains an element $b \neq 0$ such that $S[b^{-1}]$ is a field, then S is a field.

Proof. a $\Rightarrow$ b: Since R is Jacobson, S is Jacobson, and since S is a domain, it follows that the intersection of the maximal ideals of S is 0. The primes of $S[b^{-1}]$ correspond to the primes of S that do not contain b, as illustrated in the beginning of Chapter 2. Since $S[b^{-1}]$ is a field, b is contained in all the nonzero prime ideals that S may have. Thus the ideal (0) must be a maximal ideal—that is, S is a field.

b $\Rightarrow$ a: Let Q be a prime ideal of R, and let I be the intersection of all the maximal ideals containing Q. We must show that $I = Q$. If, on the contrary, $I \neq Q$, choose an element $f \in I - Q$. By Zorn's lemma we may choose a prime P maximal among the primes of R containing Q but not containing f. By hypothesis, P is not a maximal ideal of R, so $S = R/P$ is not a field. Nevertheless, P generates a maximal ideal of $R[f^{-1}]$, so $R/P[f^{-1}]$ is a field. This contradicts hypothesis b, so $I = Q$ as required. □

Proof of Theorem 4.19. We begin with an easy special case. Suppose that R is a field, and $S = R[x]$, the polynomial ring in one variable over R. The ring S is a principal ideal domain. Any nonzero prime ideal $\mathfrak{n}$ of S is generated by an irreducible monic polynomial f. Since one irreducible polynomial cannot divide another, $\mathfrak{n}$ is a maximal ideal. Of course $\mathfrak{n} \cap R = 0$, the unique maximal ideal of R, since otherwise $\mathfrak{n}$ would not be proper. The

dimension of $S/\mathfrak{n}$ over R is equal to the degree of f, and is in particular finite. Thus the second statement of Theorem 4.19 is satisfied.

It now suffices to show that S is Jacobson, and since the nonzero primes are maximal it only remains to show that 0 is the intersection of prime ideals of S. Since no polynomial can have infinitely many irreducible factors it suffices to show that S has infinitely many distinct prime ideals. For this we may use Euclid's famous old argument: If there were only finitely many prime polynomials f_i then $\prod_i f_i + 1$ (which is not a unit because it has positive degree) would have no prime factors. Thus S is a Jacobson ring, and we have proved the special case of Theorem 4.19.

Now let R be any Jacobson ring, and suppose that S is generated as an R-algebra by just one element. For the first statement, we use the characterization of Lemma 4.20, and we must show that if P is a prime of S and if $S' := S/P$ contains an element b such that $S'[b^{-1}]$ is a field, then S' is a field. Replacing S by S', and factoring out the preimage of P from R, we may assume that R is a domain contained in S, and that $b \in S$ is such that $S[b^{-1}]$ is a field, and we must show that S is a field. We shall actually show that R is also a field, and S is a finite extension of R in this case. For the second statement of the theorem, we may make the same reduction and assume that S itself is a field. The desired conclusion is exactly that R is a field and S is finite over it, so the same proof will prove both statements.

Since S is generated over R by a single element t, we may write $S = R[x]/Q$ for some prime ideal Q of $R[x]$, in such a way that t is the image of x. We first claim that $Q \neq 0$. In the contrary case, we would have $b \in R[x]$ such that $R[x][b^{-1}]$ is a field. If we write K for the quotient field of R, then $K[x][b^{-1}]$ would of course also be a field. Since we already know that $K[x]$ is Jacobson, this contradicts Lemma 4.20. Thus $Q \neq 0$, and $S[b^{-1}] = K[x]/QK[x]$ is a field, finite dimensional over K.

Let $p(x) \in Q$ be a nonzero polynomial with coefficients in R, so that

$$p(t) = p_n t^n + \cdots + p_0 = 0$$

in S. If we invert p_n, then we can multiply $p(t)$ by p_n^{-1} and we see that $S[p_n^{-1}]$ is integral over $R[p_n^{-1}]$. The element b will satisfy an algebraic equation with coefficients in R too, say

$$q(b) = q_m b^m + \cdots + q_0 = 0.$$

Since S is a domain, we may divide by a power of b if necessary and assume that $q_0 \neq 0$. Multiplying q by $1/(q_0 b^m)$ and writing β for b^{-1}, produces

$$\beta^m + (q_1/q_0)\beta^{m-1} + \cdots + (q_m/q_0) = 0.$$

Thus the field $S[\beta]$ is integral over the ring $R[(p_n q_0)^{-1}]$. By Corollary 4.17, $R[(p_n q_0)^{-1}]$ is a field. Since R is a Jacobson ring, R itself is a field. Thus S is integral over R. Again by Corollary 4.17, S is a field. This completes the proof of Theorem 4.19 in the case where S is generated by one element.

The general case may be done by induction on the number of generators r of S as an R-algebra. We may suppose that $r > 1$ and that the result has

been proved for algebras with $\leq r-1$ generators. Let S' be the subalgebra of S generated by $r-1$ of the generators of S. By induction S' is a Jacobson ring, so, by the case $r = 1$, S is a Jacobson ring too. Similarly, if $\mathfrak{n}$ is a maximal ideal of S, then $S' \cap \mathfrak{n}$ is a maximal ideal by the case $r = 1$, and $R \cap \mathfrak{n} = R \cap (S' \cap \mathfrak{n})$ is maximal by the induction step. Since the extensions $R/(R \cap \mathfrak{n}) \subset S'/(S' \cap \mathfrak{n})$ and $S'/(S' \cap \mathfrak{n}) \subset S/\mathfrak{n}$ are finite by the inductive hypothesis, $R/(R \cap \mathfrak{n}) \subset S/\mathfrak{n}$ is finite, completing the proof. □

Before proving the version of the Nullstellensatz from Chapter 1, it is convenient to prove Corollary 1.9. For convenience, we recall the statements.

Corollary 1.9. *Let k be a field. For each $p = (a_1, \ldots, a_r) \in \mathbf{A}^r(k)$ the ideal $\mathfrak{m}_p := (x_1 - a_1, \ldots, x_n - a_n) \subset k[x_1, \ldots, x_r]$ is a maximal ideal. If k is algebraically closed and $X \subset \mathbf{A}^r(k)$ is an algebraic set, then every maximal ideal of $A(X)$ is of the form $\mathfrak{m}_p/I(X)$ for some $p \in X$. In particular, the points of X are in one-to-one correspondence with the maximal ideals of the ring $A(X)$.*

Proof. It is clear that $k[x_1, \ldots, x_r]/\mathfrak{m}_p = k$, so $\mathfrak{m}_p$ is a maximal ideal. The natural map $k[x_1, \ldots, x_r] \to k[x_1, \ldots, x_r]/\mathfrak{m}_p = k$ may be described as evaluation at p. Thus $\mathfrak{m}_p \supset I(X)$ iff $p \in X$. Since the maximal ideals of $A(X)$ are the maximal ideals of $S := k[x_1, \ldots, x_r]$ containing $I(X)$, taken modulo $I(X)$, it only remains to show that every maximal ideal of S has the form $\mathfrak{m}_p$ for some p.

Suppose $\mathfrak{n}$ is a maximal ideal of S. By Theorem 4.19 applied with $R = k$, $S/\mathfrak{n}$ is algebraic over $k/(\mathfrak{n} \cap k) = k$. Since k is algebraically closed, $S/\mathfrak{n} = k$. Let a_i be the image of x_i under the map $S \to S/\mathfrak{n} = k$, and let $p = (a_1, \ldots, a_r)$. It follows that $\mathfrak{m}_p$ is contained in $\mathfrak{n}$. Since $\mathfrak{m}_p$ is maximal, $\mathfrak{m}_p = \mathfrak{n}$. □

Theorem 1.6. *Let k be an algebraically closed field. If $I \subset k[x_1, \ldots, x_n]$ is an ideal, then*

$$I(Z(I)) = \operatorname{rad} I.$$

Thus the correspondences $I \mapsto Z(I)$ and $X \mapsto I(X)$ induce a bijection between the collection of algebraic subsets of $\mathbf{A}^n(k)$ and radical ideals of $k[x_1, \ldots, x_n]$.

Proof. Corollary 1.9 shows that the points of $Z(I)$ correspond to the maximal ideals of $k[x_1, \ldots, x_n]$ containing I. Thus $I(Z(I))$ is the intersection of all the maximal ideals containing I. By Theorem 4.19, this is the same as the intersection of all the prime ideals containing I, which is $\operatorname{rad} I$ by Corollary 2.12. Since the equality $Z(I(X)) = X$ is automatic for an algebraic set X, the last statement follows. □

4.6 Exercises

Exercise 4.1: Let R be a ring, S an R-algebra, and M an S-module. The structure map from R to S makes M an R-module. If S is finite over R, and M is finitely generated as an S-module, show that M is finitely generated as an R-module.

Exercise 4.2: Let R be a domain containing a polynomial ring in one variable over a field, say $R \supset S = k[t]$. Show that if R is a finitely generated S-module, then R is free as an S-module. Show by giving a basis that if $R = k[x, y]/(x^2 - y^3)$ and $t = x^m y^n$, then the rank of R as an S-module is $3m + 2n$. Assuming again only that R is a finitely generated $S = k[t]$-module, let $\bar{R}$ be the integral closure of R. By Noether's theorem 4.14, $\bar{R}$ is again finitely generated and thus free as an S-module. Show that it has the same rank as R.

Exercise 4.3: Suppose that S is a ring, and that an element $s \in S$ satisfies an equation

$$r_0 s^n + r_1 s^{n-1} + \cdots + r_n = 0$$

with coefficients $r_i \in S$. Show that if r_n is a unit then s is a unit. Use this to deduce Corollary 4.17 without using Lemma 4.16; and then derive the latter from the former.

Exercise 4.4:* Let k be a field and let $R = k[t]/(t^2)$. Set

$$p(x) = tx^3 + tx^2 - x^2 - x \in R[x].$$

Show that $S = R[x]/(p)$ is a free R-module of rank 2, even though p is not monic (its leading coefficient is not even a unit). How do you reconcile this with Proposition 4.1?

Nakayama's Lemma

Exercise 4.5: Let $R = k[x]_{(x)}$ be the ring of polynomials in one variable x over a field, localized at the prime (x). Find an R-module M that is not finitely generated but such that M/xM is finitely generated.

Exercise 4.6: Here are two cases where Nakayama's lemma works without the finiteness condition; a third will be found in Exercise 7.2: Let R be a ring, I an ideal, and M an R-module such that $IM = M$:

a. If R is graded (by the positive integers), I is homogeneous and consists of elements of strictly positive degree, and M is a graded module with $M_n = 0$ for $n \ll 0$, show that $M = 0$.

b. If I is nilpotent, show that $M = 0$.

Exercise 4.7: Show that the Jacobson radical of R is

$$J = \{r \in R | 1 + rs \text{ is a unit for every } s \in R\}.$$

Exercise 4.8: Give a proof of Nakayama's lemma, 4.8a, without using determinants, as follows: Do induction on the number of generators required for M, and use the equation $M = IM$ to write a unit times one of the generators in terms of the others.

Exercise 4.9: The following is valid either in a local ring (R, P) or a positively graded ring R such that R_0 is local (in which case we take P to be the maximal homogeneous ideal).

Let $I \subset R$ be an ideal, and suppose $x \in P$ is an element such that x is a nonzerodivisor on R/I. Show that any minimal set of generators for I reduces mod x to a minimal set of generators for the image of I in $R/(x)$. Show by example that this can fail if x is a zerodivisor on R/I.

Exercise 4.10: Give a proof of the assertions of Corollary 4.4a and b in the special case where M is free of finite rank by showing that the nth exterior power of a surjection is a surjection.

Projective Modules and Locally Free Modules

Exercise 4.11:* Let R be a ring. **Projective** modules over R are defined in section A3.3. One description (Proposition A3.1) is that an R-module is projective iff it is a direct summand of a free R-module.

a. Use Nakayama's lemma to show that if R is local and M is a finitely generated projective module, then M is free. (This is also true without finite generation; see Kaplansky [1958] or Lam [1978].) If R is a positively graded ring, with R_0 a field, and M is a finitely generated graded module that is projective, then M is a graded free module (that is, a direct sum of the form $\oplus R(a_i)$ for some integers a_i.)

b. Use Proposition 2.10 to show that a finitely presented module M is projective iff M is locally free in the sense that the localization M_P is free over R_P for every maximal ideal of R (and then of course M_P is free over R_P for every prime ideal P of R).

Exercise 4.12:

a.* Show that if M is a finitely presented R-module, then M is projective iff M is locally free in the stronger sense that there is a finite set of elements $f_1, \ldots, f_n \in R$ such that $(f_1, \ldots, f_n) = R$, and $M[f_i^{-1}]$ is a free $R[f_i^{-1}]$-module for every i. (Hint: Here is a useful intermediate

step: Show that if M and N are finitely presented R-modules, and $M_P \cong N_P$ for some prime P, then there is an element $f \in R$, $f \notin P$ such that $M[f^{-1}] \cong N[f^{-1}]$.)

b. Any projective module is flat. Show that if R is an integral domain but not a field, then the quotient field $K(R)$ is a flat R-module that is not projective by showing that the only map $K(R) \to R$ is 0. (In Corollary 6.6 we shall show that every finitely presented flat module is projective.)

Exercise 4.13:* A ring is called **semilocal** if it has only finitely many maximal ideals. Prove that if R is a semilocal ring and M, N are finitely presented R-modules such that $M_P \cong N_P$ for every maximal ideal P of R, then $M \cong N$. (Hint: Use Proposition 2.10 to produce maps, combine them using coefficients selected according to prime avoidance, Lemma 3.3, and prove that the result is an isomorphism using Corollaries 4.4 and 2.9.) This result can fail for rings with infinitely many maximal ideals, such as $Z[\sqrt{-5}]$ or $k[x,y]/(x^3+y^3-1)$; we shall see in Chapter 11 that, more generally, if I is an ideal in a Dedekind domain R then $I_P \cong R_P$ for every prime p of R, but we may have $I \ncong R$.

Integral Closure of Ideals

Exercise 4.14: If R is a domain and I is an ideal of R, we define the integral closure of I in R to be the set of elements $s \in R$ satisfying an equation of the form

$$s^n + r_1 s^{n-1} + \cdots + r_n = 0$$

with $r_j \in I^j$, the jth power of I, for each j. Show that s is integral over I iff there is a finitely generated R-module N, not annihilated by any element of R, such that

$$sN \subset IN.$$

Use this to show that the integral closure of I in R is an ideal.

Exercise 4.15: If R is a domain, show that every radical ideal is integrally closed in R. Show that in a principal ideal domain, every ideal is integrally closed.

Exercise 4.16: Let R be a domain, K its quotient field, and $\bar{R}$ the integral closure of R in K. If $I \subset R$ is an ideal, show that $I\bar{R} \cap R$ is contained in the integral closure of I in R. (In general, the integral closure of I in R is bigger; we shall see examples with ideals of monomials, below.)

Normalization

Exercise 4.17:* Let $R \subset S$ be rings. Show that R is integrally closed in S iff $R[x]$ is integrally closed in $S[x]$.

Exercise 4.18:* Let R be a domain. Show that R is normal iff $R[x]$ is normal.

Exercise 4.19: This exercise extends the elementary result that a monic polynomial over $\mathbf{Z}$ that can be factored over $\mathbf{Q}$ can already be factored over $\mathbf{Z}$: Let $R \subset S$ be rings with R integrally closed in S. Suppose that $h(x)$ is a polynomial in $R[x]$ that factors in $S[x]$ as the product of two monic polynomials $h(x) = f(x)g(x)$. Show that f and g are each in $R[x]$. (This result leads to a solution of Exercise 4.17 different than the one given in the hint. See Atiyah and Macdonald [1969, Chapter 5, Exercise 8–9].)

Exercise 4.20: For each $n \in \mathbf{Z}$, find the integral closure of $\mathbf{Z}[\sqrt{n}]$ as follows:

a. Reduce to the case where n is square-free.

b. $\sqrt{n}$ is integral, so what we want is the integral closure R of $\mathbf{Z}$ in the field $\mathbf{Q}[\sqrt{n}]$. If $\alpha = a + b\sqrt{n}$ with $a, b \in \mathbf{Q}$, then the minimal polynomial of α is $x^2 - \text{Trace}(\alpha)x + \text{Norm}(\alpha)$, where $\text{Trace}(\alpha) = 2a$ and $\text{Norm}(\alpha) = a^2 - b^2 n$. Thus $\alpha \in R$ iff $\text{Trace}(\alpha)$ and $\text{Norm}(\alpha)$ are integers.

c. Show that if $\alpha \in R$ then $a \in {}^1\!/\!_2\mathbf{Z}$. If $a = 0$, show $\alpha \in R$ iff $b \in \mathbf{Z}$. If $a = {}^1\!/\!_2$ and $\alpha \in R$, show that $b \in {}^1\!/\!_2\mathbf{Z}$. Thus, subtracting a multiple of $\sqrt{n}$, we may assume $b = 0$ or ${}^1\!/\!_2$. $b = 0$ is impossible.

d. Conclude that the integral closure is $\mathbf{Z}[\sqrt{n}]$ if $n \not\equiv 1 (\text{mod}\, 4)$, and is $\mathbf{Z}[{}^1\!/\!_2 + {}^1\!/\!_2\sqrt{n}]$ if $n \equiv 1 (\text{mod}\, 4)$.

Exercise 4.21 (The graded case):

a.* Show that the integral closure of a graded domain in its quotient field is graded, as follows: First, the degree 0 part of the graded ring obtained by inverting all nonzero homogeneous elements of S is a field. Next, show that a domain S is normal iff the ring of Laurent polynomials $S[x, x^{-1}]$ is normal. Finally, show that if $S \subset T$ are graded domains (T can even be $\mathbf{Z}$-graded), with S Noetherian, then the integral closure of S in T is again graded.

b. If S is a graded Noetherian domain, show that for any homogeneous prime ideal P of S not containing S_1, the integral closure of $S_{(P)}$ is the degree 0 part of a localization of the integral closure of S.

Normalization and Convexity

The operation of normalizing has many similarities to the operation of taking convex hulls, and indeed there is more than an analogy between these ideas. Here are two cases where the correspondence is very tight.

Exercise 4.22: Let $\Gamma \subset \mathbf{N}^n$ be a finitely generated subsemigroup (with identity) of the nth power of the semigroup of natural numbers under addition. Let k be a field, and define

$$k[\Gamma] \subset k[x_1, \ldots, x_n]$$

to be the subring that is spanned as a vector space by all the monomials with exponents in Γ; that is, by all

$$x^\gamma := x_1^{\gamma_1} x_2^{\gamma_2} \ldots x_n^{\gamma_n} \quad \text{with} \quad \gamma = (\gamma_1, \ldots, \gamma_n) \in \Gamma.$$

We define $\mathbf{R}_+\Gamma$ to be the **convex cone** spanned by Γ; that is, $\mathbf{R}_+\Gamma$ is the set of all positive real linear combinations of elements of Γ. We define $G(\Gamma) \subset \mathbf{Z}^n$ to be the group generated by Γ. Let

$$\bar{\Gamma} = [\mathbf{R}_+\Gamma] \cap \mathbf{N}^n = [\mathbf{R}_+\Gamma] \cap G(\Gamma),$$

the semigroup of all integral points in the cone spanned by Γ. Show that $k[\bar{\Gamma}]$ is the integral closure of $k[\Gamma]$ in its quotient field as follows:

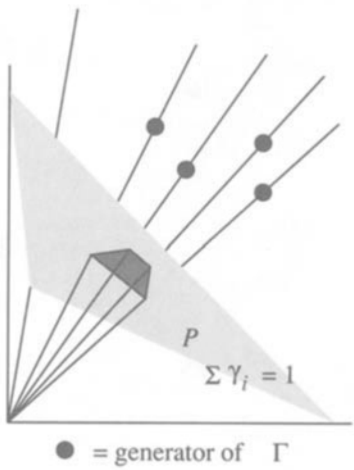
a. Show that

$$\bar{\Gamma} = \{\gamma \in \mathbf{N}^n | m\gamma \in \Gamma \text{ for some positive integer } m.\}$$

To do this, first prove

Caratheodory's Theorem. *Let $v_1, \ldots, v_m \in \mathbf{R}^n$ be points. The set $\mathbf{R}_+v_1 + \cdots + \mathbf{R}_+v_m$ is the union of the sets $\mathbf{R}_+v_{i_1} + \cdots + \mathbf{R}_+v_{i_s}$ where $v_{i_1}, \ldots, v_{i_s}$ are linearly independent.*

Use this to show that any rational point of $\mathbf{R}_+\Gamma$ is a positive rational linear combination of elements of Γ.



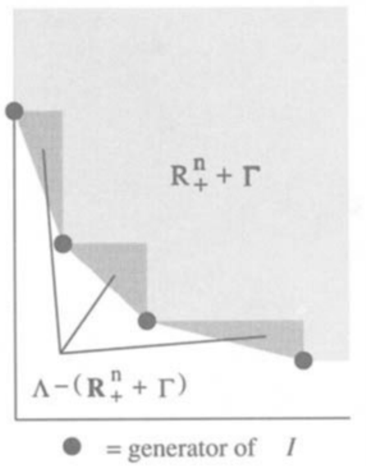
b. Show that any monomial in $k[\bar{\Gamma}]$ is integral over $k[\Gamma]$: Make an integral equation from a relation of the form $m\alpha = \gamma \in \Gamma$ with $m \in \mathbf{N}$.

c. Given any element f of the quotient field of $k[\Gamma]$, integral over $k[\Gamma]$, show that $f \in k[G(\Gamma)] \cap k[\mathbf{N}^n]$: In fact, $G(\Gamma)$ is a free abelian group, so both the rings $k[G(\Gamma)]$ and $k[\mathbf{N}^n]$ are normal. Thus any element of the integral closure of $k[\Gamma]$ can be written as a polynomial f whose monomials lie in $k[G(\Gamma)]$. It remains to show that all the monomials of f lie in $\mathbf{R}_+\Gamma$.

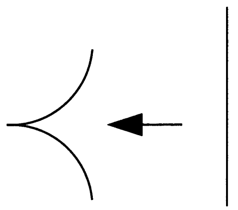
d. Let Γ' be the semigroup generated by Γ and the monomials of f. Let P be the set $\{\gamma_1, \ldots, \gamma_n) \in \mathbf{R}^n | \sum_i \gamma_i = 1\}$, as in the figure. If not all the exponents of monomials of f are in Γ, then one of these exponents, say α, lies on the ray through an extremal vertex of the convex set $\mathbf{R}_+\Gamma \cap P$. Thus we may find a linear functional L on $\mathbf{R}^n$ with value > 0 on α and < 0 on all the other exponents of monomials in f and on all monomials in Γ.

e. Let $f^n + a_{n-1}f^{n-1} + \cdots + a_n = 0$ be the integral equation satisfied by f, with all the $a_i \in k[\Gamma]$. Evaluating L on the exponents of monomials that occur in each term on the left-hand side, we find that the maximum value is taken on only at $n\alpha$, and thus $x^{n\alpha}$ occurs only once. Thus the left hand side cannot be 0 as claimed. The contradiction shows that $f \in k[\bar{\Gamma}]$.

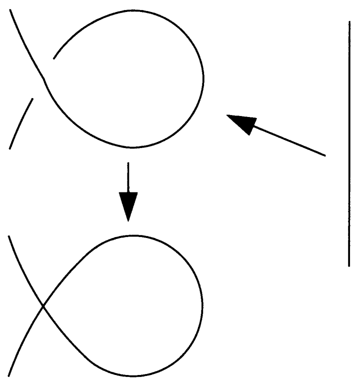
Exercise 4.23: Let $I \subset k[x_1, \ldots, x_n]$ be an ideal generated by monomials in a polynomial ring over a field k. Let Γ be the set of exponents of monomials in I, so that I is the linear span of the monomials x^γ for $\gamma \in \Gamma$. Regarding Γ as a subset of $\mathbf{N}^n \subset \mathbf{R}^n_+$, we let Λ be the convex hull of $\mathbf{R}^n_+ + \Gamma$, as in the figure, and we let Γ^* be the set of integral points in Λ. Show that the integral closure of I is the ideal generated by Γ^*.



Exercise 4.24: Let R be either of the domains $\mathbf{C}[x,y]/(y^2 - x^3)$, or $\mathbf{C}[x,y]/(y^2 - x^2(x+1))$, and let $t = y/x$, an element of the quotient field. Show that in each case, $R[t] = \mathbf{C}[t]$. Pictorially, the normalization maps are as follows:



and



Exercise 4.25: Let X be an affine variety over $\mathbf{C}$, and let $R = \mathbf{C}[x_1, \ldots, x_n]/I$ be its coordinate ring. Show that if $p(x)/q(x)$ is an element of the quotient field of R that is integral over R, then for each point $x \in X$ there is a neighborhood U of x and a real constant B such that the absolute value $|p(x)/q(x)|$ is bounded by B at all the points of U where q is nonzero. (The converse is also true, but requires a deeper characterization of the integral closure.)

Nullstellensatz

Exercise 4.26: Suppose that the additive group of the ring R is a finitely generated abelian group. If P is a maximal ideal of R, show that R/P is a finite field. Show that every prime ideal of R that is not maximal is a minimal prime ideal.

Exercise 4.27 (Maximal ideals of a polynomial ring): Let k be a field and let $R = k[x_1, \ldots, x_r]$ be a polynomial ring. Show that any maximal ideal of R may be generated by r elements f_i, where f_i is a polynomial depending only on $x_1, \ldots, x_i$. See also Exercise 13.6.

Exercise 4.28: Suppose that k is an algebraically closed field, and let $\mathfrak{m}$ be a maximal ideal of the polynomial ring $R = k[x_1, \dots, x_r]$. Show that there is a k-algebra automorphism of R taking $\mathfrak{m}$ to the ideal $(x_1, \dots, x_r)$.

Three More Proofs of the Nullstellensatz

We now give three further proofs of the form of the Nullstellensatz given in Theorem 1.6. The following provides what I think is the fastest proof of the Nullstellensatz in case the ground field is $\mathbf{C}$.

Exercise 4.29 (Quick and dirty proof of the Nullstellensatz): Let K be an algebraically closed field of infinite transcendence degree over a prime field $k(= \mathbf{Q}$ or $= \mathbf{Z}/(p))$, such as $\mathbf{C}$.

Show that each prime P of $K[x_1, \dots, x_r]$ is the intersection of maximal ideals of the form $\mathfrak{m}_p = (x_1 - a_1, \dots, x_r - a_r)$ by showing that given any $f \notin P$, there is a point $p \in Z(P)$ such that $f(p) \neq 0$ (this proves Corollary 1.9 immediately, and with Corollary 2.12 it proves Theorem 1.6) as follows:

Show that there are elements $\alpha_1, \dots, \alpha_n \in K$ such that writing K' for $k(\alpha_1, \dots, \alpha_n)$, and $P' = P \cap K'[x_1, \dots, x_r]$, we have $P = P'K[x_1, \dots, x_r]$. Show that P' is a prime ideal.

Show that the quotient field L of $K'[x_1, \dots, x_r]/P'$ may be embedded in K. (This is field theory: First embed the transcendental part, then the algebraic part. This is where you need the transcendence degree of K to be large—$\geq r$ would do.) Let a_i be the image of x_i in K under this embedding.

Now show that $p = (a_1, \dots, a_r)$ has the desired property.

(Remark for those who know some model theory: The only drawback of this amazingly easy proof is that it requires K to have large transcendence degree. This may be overcome by using some logic. If we start with any field K, we may begin by replacing it with a nontrivial ultrapower to get an "elementarily equivalent" field of infinite transcendence degree (see, for example, Bell and Slomson [1969] for unexplained terminology). The truth of the statement we need to prove can be transferred from this ultrapower back to K.)

The next two proofs of the Nullstellensatz depend on a classic localization argument that is another version of "Rabinowitch's trick".

Exercise 4.30: Suppose that k is a Noetherian ring such that

*) for every finitely generated k-algebra R and maximal ideal $P \subset R$ the k-algebra R/P is finite over k.

Show that for every reduced finitely generated k-algebra R and prime ideal $Q \subset R$ we have $Q = \cap P$, where the intersection runs over all primes P of R such that R/P is finite over k. (Hint: If $f \in R$, $f \notin Q$, we must find a

prime P such that R/P is finite over k and $f \notin P$. Consider a maximal ideal in the k-algebra $R[f^{-1}]$ and its intersection with R.)

Deduce in particular that Theorem 1.6 (for a given field k) follows if we prove that $*)$ holds for k.

Exercise 4.31 (Nullstellensatz for uncountable fields): The following simple argument of Krull and Van der Waerden is the fastest way to check the hypothesis $*)$ of Exercise 4.30 for the complex numbers. Unfortunately, it works only for uncountable fields.

Show that if k is a field and $k(x)$ is the field of rational functions in one variable over k, then the elements of the set $\{1/(x-a)|a \in k\} \subset k(x)$ are linearly independent, so $\dim_k k(x) \geq \operatorname{card} k$. Deduce that if K is an extension field of k, and $\dim_k K < \operatorname{card} k$, then K is algebraic over k.

On the other hand, show that if R is a finitely generated k-algebra, then $\dim_k R$ is at most countable.

Deduce that hypothesis $*)$ of Exercise 4.30, and thus also Theorem 1.6, holds for uncountable fields.

Exercise 4.32 (Artin-Tate Proof of the Nullstellensatz): E. Artin and J. Tate in [1951] (reprinted in Artin [1965]) found a remarkable result that implies hypothesis $*)$ of Exercise 4.30 for any field:

Theorem (Artin-Tate). *Suppose R is a Noetherian ring and S is a finitely generated R-algebra. If $T \subset S$ is an R-algebra such that S is a finitely generated T-module, then T is a finitely generated R-algebra.*

Prove the Artin-Tate Theorem. Suppose $x_1, \ldots, x_r$ generate S as an R-algebra. The fact that just finitely many elements $s_1, \ldots, s_q$ generate S as a T-module can be written down in terms of finitely many elements of T as follows. There exist elements t_{ij} and t'_{ijk} in T such that

$$x_i = \sum t_{ij} s_j$$
$$s_i s_j = \sum t'_{ijk} s_k.$$

Let T_0 be the subalgebra of T generated over R by the t_{ij} and the t'_{ijk}. Show that S is finitely generated as a T_0-module. Conclude that T is finitely generated as a T_0-module. Use the fact that T_0 is finitely generated over R to finish the proof of the theorem.

Deduce that hypothesis $*)$ of Exercise 4.30 holds for any field k:

1. Show that the fields that could appear as R/P in the statement are exactly those extension fields of k that are finitely generated as algebras. Thus, $*)$ amounts to saying that if K is a field that is finitely generated as a k-algebra, then K is finitely generated as a k-module (that is, K is a finite field extension of k).

2. Note that it is enough to prove that K is algebraic over k. Suppose that $x_1, \ldots, x_r \in K$ is a transcendence base for K over k. Use the Artin-Tate theorem to show that the field of rational functions $k(x_1, \ldots, x_r)$ is a finitely generated k-algebra. Show that this implies that $k(x_1, \ldots, x_r) = k[x_1, \ldots, x_r][f^{-1}]$ for some polynomial f. Conclude that every prime of $k[x_1, \ldots, x_r]$ must divide f. But if $r > 0$, the following exercise shows that there are infinitely many primes, so $r = 0$.

Exercise 4.33: Let k be a field. Exhibit infinitely many maximal ideals of $k[x]$. If k is infinite, show that there are infinitely many maximal ideals with residue field k; if k is finite, one must consider finite extension fields of k.

5

Filtrations and the Artin-Rees Lemma

In this chapter we shall describe two constructions—the associated graded ring and the blowup algebra—that are made from a **descending multiplicative filtration** of a ring R; that is, from a sequence of ideals

$$R = I_0 \supset I_1 \supset I_2 \supset \cdots \text{ satisfying } I_i I_j \subset I_{i+j} \quad \text{for all } i, j.$$

A third such construction, the Rees algebra, is treated at the end of the next chapter, and sheds some light on the results we shall prove about the associated graded ring. Chapter 7 will be devoted to a fourth example, the completion. Each is used to get information about R by comparing it with a closely related ring that is simpler in some way.

These constructions are most often used in the case where the I_j are the powers of a single ideal, $I_j = I^j$; this is called the ***I*-adic filtration**. In applications I is often taken to be the maximal ideal of a local Noetherian ring R, and the reader will not lose too much by imagining this to be the case throughout.

It is quite useful to generalize to the case of modules, so we also study the I-adic filtration of a module M, say $M \supset IM \supset I^2M \supset \cdots$. But if we intersect the terms of the I-adic filtration of M with some submodule $M' \subset M$, we do not generally get the I-adic filtration of M'. A key result of the theory, the Artin-Rees lemma[1], shows that the induced filtration is often **stable** in the following sense:

---

[1]The paper of Rees [1956] contains a special case. According to Nagata, who seems to have coined the name, Artin lectured on the general case in 1955.

Definitions. *Let R be a ring, let $I \subset R$ be an ideal, and let M be an R-module. A filtration $M = M_0 \supset M_1 \supset \cdots$ is called an **I-filtration** if $IM_n \subset M_{n+1}$ for all $n \geq 0$. An I-filtration is called **I-stable** if in addition $IM_n = M_{n+1}$ for all sufficiently large i. When the ideal I is understood, we speak simply of **stable filtrations**.*

An I-stable filtration is determined if one knows a sufficiently large finite number of the M_i; in this sense it is a finitely generated filtration.

Lemma 5.1 (Artin-Rees). *Let R be a Noetherian ring, let $I \subset R$ be an ideal, and let $M' \subset M$ be finitely generated R-modules. If $M = M_0 \supset M_1 \supset \cdots$ is an I-stable filtration, then the induced filtration $M' \supset M' \cap M_1 \supset M' \cap M_2 \supset \cdots$ is also I-stable. That is, there exists a number n such that for all $i \geq 0, M' \cap M_{i+n} = I^i(M' \cap M_n)$.*

We shall give the proof later in this chapter after having defined some basic constructions. For an interesting recent development in this theory, see Huneke [1992].

5.1 Associated Graded Rings and Modules

Let I be an ideal of a ring R. We define the associated graded ring of R with respect to I, written $\operatorname{gr}_I R$ to be the graded ring

$$\operatorname{gr}_I R := R/I \oplus I/I^2 \oplus \cdots .$$

Here the multiplication in $\operatorname{gr}_I R$ is given as follows: If $a \in I^m/I^{m+1}$ and $b \in I^n/I^{n+1}$, then taking representatives a' and b' of a and b in I^m and I^n, respectively, we define $ab \in I^{m+n}/I^{m+n+1}$ to be the image of $a'b'$. Note that this is well-defined modulo I^{m+n+1}.

More generally, let $\mathfrak{I} : M = M_0 \supset M_1 \supset \cdots$ be an I-filtration of an R-module M. Let

$$\operatorname{gr}_{\mathfrak{I}} M := M/M_1 \oplus M_1/M_2 \oplus \cdots .$$

We make $\operatorname{gr}_{\mathfrak{I}} M$ into a graded $\operatorname{gr}_I R$-module as follows: If $a \in I^m/I^{m+1}$ and $b \in M_n/M_{n+1}$ have representatives $a' \in I^m$ and $b' \in M_n$, then ab is the class of $a'b'$ in M_{m+n}/M_{m+n+1}. The assumption that $\mathfrak{I}$ is an I-filtration ensures that this is well defined. When there is no danger of confusion, we shall simply write $\operatorname{gr} M$ for $\operatorname{gr}_{\mathfrak{I}} M$.

The following elementary result explains the importance of the stability property:

Proposition 5.2. *Let I be an ideal in a ring R, and suppose that M is a finitely generated R-module. If $\mathfrak{I} : M = M_0 \supset M_1 \supset \cdots$ is an I-stable*

filtration by finitely generated submodules of M, then $\operatorname{gr}_{\mathcal{J}} M$ is a finitely generated module over $\operatorname{gr}_I R$.

Proof. Suppose that $IM_i = M_{i+1}$ for all $i \geq n$. Clearly, $(I/I^2)(M_i/M_{i+1}) = M_{i+1}/M_{i+2}$ for $i \geq n$. Thus the union of any sets of generators of the modules $M_0/M_1, \ldots, M_n/M_{n+1}$ will generate $\operatorname{gr} M$. Since each M_i is finitely generated, each of these sets of generators may be chosen to be finite. □

See Appendix 3, particularly Theorem A3.22 and Exercise A3.42, for another use of stability.

Let M be an R-module with filtration $\mathcal{J} : M = M_0 \supset M_1 \supset \cdots$. There is no interesting natural homomorphism from M to $\operatorname{gr} M$, but there is an interesting natural map of sets defined as follows: Given $f \in M$, let m be the greatest number such that $f \in M_m$, and define the **initial form of f**, denoted **in(f)** by

$$\operatorname{in}(f) = f \text{ modulo } M_{m+1} \in M_m/M_{m+1} \subset \operatorname{gr} M,$$

or by

$$\operatorname{in}(f) = 0 \text{ if } f \in \cap_1^\infty M_m.$$

See Exercise 5.1 for some of the properties of this map.

Now suppose that $I \subset R$ is an ideal that $\mathcal{J}$ is an I-filtration of M. Set $G = \operatorname{gr}_I R$. If $M' \subset M$ is a submodule, we define $\operatorname{in}(M')$ to be the G-submodule of $\operatorname{gr} M$ generated by $\operatorname{in}(f)$ for all $f \in M'$. The submodule $\operatorname{in}(M')$ is generally *not* obtained as the submodule generated by the initial forms of a given set of generators of M'. For example, if

$$J = (xy + y^3, x^2) \subset R = k[x, y],$$

and $I = (x, y)$, then with respect to the I-adic filtration, $\operatorname{in}(x^2) = x^2$ and $\operatorname{in}(xy + y^3) = xy$ but x^2 and xy do not generate $\operatorname{in}(J)$. For example,

$$x(xy + y^3) - y(x^2) = xy^3 \in J,$$

so

$$y^2(xy + y^3) - xy^3 = y^5 \in J,$$

and thus $y^5 \in \operatorname{in}(J)$. In fact, $\operatorname{in}(J) = (x^2, xy, y^5)$. In Chapter 15 we shall exhibit a general technique for handling such computations. See Exercise 5.2 for an easy special case.

A first hint of why the associated graded construction is interesting is given by the fact that if I is a maximal ideal then $\operatorname{gr}_I R$ is a graded algebra over the field R/I. It is even a finitely generated algebra if, as in nearly all cases of interest, I is a finitely generated ideal. Thus gr_I gives us a way of turning arbitrary local Noetherian rings into finitely generated graded rings, about which we know more. For example, we get a ready-made theory of Hilbert functions:

Definition. *If R is a local ring with maximal ideal I, then the* ***Hilbert function of R*** *is the function*

$$H_R(n) = \dim_{R/I} I^n/I^{n+1}.$$

More generally, if M is an R-module, we can define

$$H_M(n) = \dim_{R/I} I^n M/I^{n+1}M.$$

Since these are just the Hilbert functions of $\operatorname{gr}_I R$ and $\operatorname{gr}_I M$, and $\operatorname{gr}_I M$ is a finitely generated $\operatorname{gr}_I R$-module, we already know that they agree for large n with polynomials $P_R(n)$ and $P_M(n)$ of degree $\leq H_R(1) - 1$, and that H_R and H_M can be expressed exactly in terms of binomial coefficients. These functions are quite important in dimension theory, and we shall return to them in Chapter 12.

One can sometimes derive nice properties of R from nice properties of $\operatorname{gr}_I R$. To do this we need to know that no elements of R have been "forgotten" by $\operatorname{gr}_I R$, as would be the case if an element of R were in every power of I. Fortunately, $\cap_j I^j = 0$ in most cases of interest. The tool we need to prove this is the Artin-Rees lemma. The proof of the lemma uses another construction of great geometric and algebraic interest, to which we now turn.

5.2 The Blowup Algebra

Definition. *If R is a ring and $I \subset R$ is an ideal, then the* ***blowup algebra of I in R*** *is the R-algebra*

$$B_I R := R \oplus I \oplus I^2 \oplus \cdots \cong R[tI] \subset R[t].$$

Note that $B_I R/IB_I R = R/I \oplus I/I^2 \oplus \cdots = \operatorname{gr}_I R$, the associated graded ring.

The geometric context in which the blowup algebra arises accounts for the name: If R is the coordinate k-algebra of an affine algebraic set X over k, and I is the ideal of an algebraic subset $Y \subset X$, then there is an algebraic set Z obtained by a process called **blowing up $Y \subset X$**, defined as follows: Let $a_1, \ldots, a_r$ be k-algebra generators for R and let $g_0, \ldots, g_s$ be generators of I as an ideal of R. The algebra $B_I R$ is a homomorphic image of the ring $k[x_1, \ldots, x_r, y_0, \ldots, y_s]$ by the map sending $x_i \mapsto a_i$ and $y_j \mapsto g_j$. The kernel of this map is an ideal that is easily seen to be homogeneous in the variables y_j. It thus corresponds to an algebraic subset $Z \subset \mathbf{A}^r \times \mathbf{P}^s$. The projection map $\mathbf{A}^r \times \mathbf{P}^s \to \mathbf{A}^r$ maps Z onto X and is an isomorphism away from the preimage of Y. The set Z is called the **blowup**[2] **of Y in X**. The preimage of Y in Z corresponds to the ring $B_I R/IB_I R = \operatorname{gr}_I R$.

[2]The name "blowup" may come from the most commonly used special case: When one blows up a point on a smooth complex surface, one replaces the point

This preimage, which is called the **exceptional set** of the blowup, is the projective variety associated to the graded ring $\operatorname{gr}_I R$.

If M is an R-module and $\mathcal{J} : M = M_0 \supset M_1 \supset \cdots$ is an I-filtration, then the direct sum

$$B_{\mathcal{J}}M := M \oplus M_1 \oplus \cdots$$

becomes a graded module over the blowup ring $B_I R$ in an obvious way. Using this construction, the connection between finitely generated modules and stable filtrations becomes even tighter.

Proposition 5.3. *Let R be a ring, let $I \subset R$ be an ideal, and let M be a finitely generated R-module with I-filtration $\mathcal{J} : M = M_0 \supset M_1 \supset \cdots$ by finitely generated modules M_i. The filtration $\mathcal{J}$ is I-stable iff the $B_I R$-module $B_{\mathcal{J}}M$ is finitely generated.*

Proof. If $B_{\mathcal{J}}M$ is finitely generated, then its generators must be contained in the direct sum of the first n terms for some n. Replacing them by their homogeneous components, we see that $B_{\mathcal{J}}M$ is generated by elements of the modules M_i for various $i \leq n$. Of course, then

$$M_n \oplus M_{n+1} \oplus \cdots$$

is generated as a $B_{\mathcal{J}}R$-module by M_n—that is,

$$M_{n+i} = I^i M_n$$

for all $i \geq 0$, so $\mathcal{J}$ is stable.

Conversely, if $\mathcal{J}$ is stable, so that $M_{n+i} = I^i M_n$ for some n and all $i \geq 0$ say, then $B_{\mathcal{J}}M$ is clearly generated by the union of any sets of generators for $M_0, M_1, \ldots, M_n$. □

With this construction, the Artin-Rees lemma becomes a corollary of the Hilbert basis theorem.

Proof of Lemma 5.1. Let

$$\mathcal{J}' : M' = M'_0 \supset M'_1 := M' \cap M_1 \supset M'_2 := M' \cap M_2 \cdots$$

be the induced filtration on M'. The module $B_{\mathcal{J}'}M'$ is naturally a graded $B_I R$-submodule of $B_{\mathcal{J}}M$. If $\mathcal{J}$ is a stable filtration, then $B_{\mathcal{J}}M$ is finitely generated by Proposition 5.3. Because $B_I R$ is a finitely generated R-algebra, it is Noetherian, so the submodule $B_{\mathcal{J}'}M'$ is finitely generated too. Proposition 5.3 now shows that $\mathcal{J}'$ is a stable filtration of M'. □

---

by a copy of $\mathbf{P}^1_{\mathbf{C}}$—topologically a 2-sphere. Topologically, this corresponds to a surgery. But I like to think of this process as that of sticking a soda straw into the surface (topologically a 4-manifold) and blowing a little bubble (2-sphere) at the point; the antipodal points of the bubble must then be identified to get back a 4-manifold. The original German word, *aufblasen*, is consistent with this interpretation. A Frenchman once suggested to me that the French translation, *éclater* (explode), was chosen *pour faire peur aux gens* (to frighten people).

5.3 The Krull Intersection Theorem

As a first application we get an important theorem of Krull [1938] (in the form given by Chevalley [1943]).

Corollary 5.4 (Krull Intersection Theorem). *Let $I \subset R$ be an ideal in a Noetherian ring R. If M is a finitely generated R-module, then there is an element $r \in I$ such that $(1-r)(\cap_1^\infty I^j M) = 0$. If R is a domain or a local ring, and I is a proper ideal, then*

$$\bigcap_1^\infty I^j = 0.$$

Proof. By the Artin-Rees lemma, applied to the submodule $\cap_1^\infty I^j M \subset M$, there is an integer p such that

$$\begin{aligned}\bigcap_1^\infty I^j M &= \left(\bigcap_1^\infty I^j M\right) \bigcap I^{p+1} M \\ &= I\left(\left(\bigcap_1^\infty I^j M\right) \cap I^p M\right) \\ &= I\left(\bigcap_1^\infty I^j M\right).\end{aligned}$$

The first statement now follows from Corollary 4.7. To prove the second statement we take $M = R$. It is enough to show that in the given cases $1-r$ is a nonzerodivisor. Since I is a proper ideal, we have at least $r \neq 1$, so $1-r \neq 0$, and if R is a domain we are done. In the case where R is local, I must be contained in the maximal ideal, so r is too. Thus $1-r$ is a unit in this case. □

A common theme, to some extent explained in the next chapter, is that good properties of $\mathrm{gr}_I R$ imply good properties for R. Here is a sample:

Corollary 5.5. *Let R be a Noetherian local ring and let I be a proper ideal of R. If $\mathrm{gr}_I R$ is a domain, then R is a domain.*

Proof. If $fg = 0$ in R, then $\mathrm{in}(f)\,\mathrm{in}(g) = 0$ in $\mathrm{gr}_I R$, so $\mathrm{in}(f)$ or $\mathrm{in}(g)$ is 0. By the Krull intersection theorem, $\cap_1^\infty I^n = 0$, so this implies that f or g is 0. □

The converse of Corollary 5.5 fails dreadfully, and quite generally $\mathrm{gr}_I R$ can be "bad" in ways in which R is "good." See Exercise 5.8.

Example. Both the Krull intersection theorem and Corollary 5.5 can fail in the non-Noetherian case. For example, let R be the ring of germs of $\mathcal{C}^\infty$ functions on $(\mathbf{R}, 0)$, and let x be the coordinate function. Let

$$g(x) = e^{-\frac{1}{x^2}} \in R,$$

whose graph is pictured in Figure 5.1. The ring R is local with maximal ideal I generated by the function x, and since $g(x)/x^n$ is $\mathcal{C}^\infty$ for every n, we see that

$$g(x) \in \bigcap_1^\infty I^n.$$

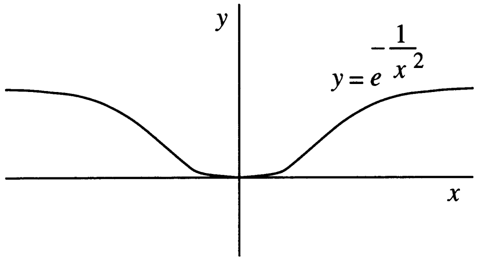


FIGURE 5.1.

In this case $\cap_1^\infty I^n$ is the set of germs of functions vanishing at the origin that are flat in the sense that all their derivatives at the origin also vanish. It is true that $I(\cap_1^\infty I^n) = \cap_1^\infty I^n$—see Exercise 5.6. But the $\mathcal{C}^\infty$ functions of the form $1 - r$ with $r \in (x)$ are just the $\mathcal{C}^\infty$ germs with the value 1 at the point 0. Thus $g(x)$, for example, is not annihilated by any such function, so Corollary 5.4 fails. For the fate of Corollary 5.5, see Exercise 5.7.

5.4 The Tangent Cone

The fact that the associated graded ring corresponds to the exceptional set in the blowup has a simple and beautiful geometric consequence. Let

$$R = k[x_1, \dots, x_r]/J, \quad I = (x_1, \dots, x_r)$$

where k is an algebraically closed field. Let $X = Z(J) \subset \mathbf{A}^r$ and suppose that $J \subset I$, so that $0 \in X$. The **tangent cone** of X at 0 is the cone composed of all lines that are the limiting positions of secant lines to X passing through the point 0. One can show that the ideal $\mathrm{in}_I(J) \subset k[x_1, \dots, x_r]$ defines the tangent cone so that the coordinate ring of the tangent cone is $\mathrm{gr}_{(x_1,\dots,x_r)} R$; see Exercise 5.3. The pictures in Figure 5.2 illustrate this point. For a proof see, for example, Harris [1992].

5.5 Exercises

Exercise 5.1: Let R be a ring and M an R-module. Suppose that $\mathcal{J} : M = M_0 \supset M_1 \supset \cdots$ is a filtration by submodules. Although the map $M \to \mathrm{gr}_{\mathcal{J}} M$ sending f to $\mathrm{in}(f)$ is not a homomorphism of abelian groups, show that either $\mathrm{in}(f) + \mathrm{in}(g) = \mathrm{in}(f+g)$ or $\mathrm{in}(f) + \mathrm{in}(g) = 0$.

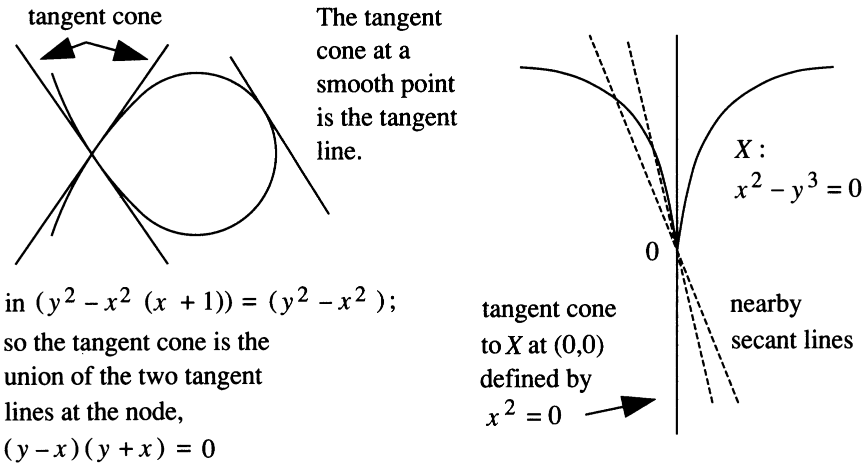


FIGURE 5.2.

Suppose that $M = R$, and that $\mathcal{J}$ is a multiplicative filtration, so that $\operatorname{gr}_{\mathcal{J}} R$ is a ring. Show that either $\operatorname{in}(f)\operatorname{in}(g) = \operatorname{in}(fg)$ or $\operatorname{in}(f)\operatorname{in}(g) = 0$.

Exercise 5.2: Let I be an ideal in a ring R. Suppose that $\operatorname{gr}_I R$ is a domain. Show that if $Rf \subset R$ is a principal ideal, then $\operatorname{in}(Rf) \subset \operatorname{gr} R$ is generated by $\operatorname{in}(f)$. Find an example of a local ring R with maximal ideal I and an element $f \in I$ such that $\operatorname{in}(Rf)$ is not generated by $\operatorname{in}(f)$.

Exercise 5.3: Suppose $J \subset I$ are ideals in a ring R. Show that

$$\operatorname{gr}_I R/J = (\operatorname{gr}_I R)/\operatorname{in}(J).$$

Exercise 5.4: Taking the associated graded ring can also simplify some features of the structure of R. For example, let k be a field, and let

$$R = k[x_1, \ldots, x_r] \subset R_1 = k[[x_1, \ldots, x_r]]$$

be the rings of polynomials in r variables and formal power series in r variables over k, and write $I = (x_1, \ldots, x_r)$ for the ideal generated by the variables in either ring. Show that

$$\operatorname{gr}_I R = \operatorname{gr}_I R_1.$$

If k is the field of real or complex numbers, write R_2 for the ring of convergent power series and R_3 for the ring of $\mathcal{C}^\infty$ functions on k^r, so that $R \subset R_2 \subset R_3$, and R_3 maps to R_1 by sending each element to its Taylor series. Denote by I the ideal generated by the variables in any of these rings. Show that

$$\operatorname{gr}_I R = \operatorname{gr}_I R_2 = \operatorname{gr}_I R_3 = \operatorname{gr}_I R_1.$$

Exercise 5.5: A converse to the Krull intersection theorem: Let $I \subset R$ be an ideal in a Noetherian ring R and let M be a finitely generated R-module. Show that there is a largest submodule $N \subset M$ such that N is annihilated by an element of the form $1 - r$ with $r \in I$. Show that $\cap_1^\infty I^j M = N$.

Exercise 5.6: Let R be the ring of germs of $\mathcal{C}^\infty$ functions on $(\mathbf{R}, 0)$, and let $I = (x)$, where x is the coordinate function. Show by elementary calculus that if f is a $\mathcal{C}^\infty$ function that vanishes with all its derivatives at the origin, then f/x is also such a function. Conclude that $I(\cap_1^\infty I^n) = \cap_1^\infty I^n$.

Exercise 5.7:* Show that the ring R of germs of $\mathcal{C}^\infty$ functions on $(\mathbf{R}, 0)$ is not a domain, although $\operatorname{gr}_I R = \mathbf{R}[x]$ is a domain.

Exercise 5.8:

a. Let $R = k[x, y]/(x^2 - y^3)$, and let $I = (x, y)$. Show that R is a domain, but $\operatorname{in}(x)^2 = 0$ in $\operatorname{gr}_I R$.

b.* Let $R = k[t^4, t^5, t^{11}] \subset k[t]$, and let $I = (t^4, t^5, t^{11})$. Show that $\operatorname{in}(I) \operatorname{in}(t^{11}) = 0$.

6
Flat Families

Recall from Chapter 2 that a module M over a ring R is **flat** if for every inclusion $N' \subset N$ of R-modules the induced map $M \otimes_R N' \to M \otimes_R N$ is again an inclusion. The notion of flatness was first isolated by Serre [1955–56] and was then systematically developed and mined by Grothendieck. It is now a central theme in algebraic geometry and commutative algebra.

We saw in Chapter 2 that the flatness of algebras of the form $R[U^{-1}]$ helps to connect their properties with the properties of R, and we shall exploit this idea again when we come to completions in Chapter 7. But flatness plays another important role as well: Flatness turns out to be a property possessed by many natural families of varieties or algebras, and it leads to good properties of these families. In this chapter we shall study flatness abstractly, but first we digress to explain the idea of flat families.

First, what is a "family" of varieties or of algebras? One has in mind a collection of objects, "varying with parameters." A typical example, essentially the first one ever considered, is the family of curves of degree d in the affine plane over a field k. Algebraically, this corresponds to the family of k-algebras $k[x, y]/(f)$, where f is a nonzero polynomial of degree d. Here the parameters are the coefficients of f. (To get the most from this example one should consider projective plane curves instead of the affine case, which we have taken for simplicity; see Exercise 6.6.)

Perhaps the most inclusive and powerful way of making the notion of family precise is to say simply that a family is a morphism! For example, if $\varphi : X \to B$ is a morphism of varieties, then the preimages of points of B, which are called the *fibers* of φ, are varieties that vary in a family parametrized by the points of B. To see the family of plane curves above

from this point of view, we take B to be the affine space of polynomials $f = \sum a_{ij}x^iy^j$ of degree d, which is an affine space of dimension $N := (d+2)(d+1)/2$, with coordinates $\{a_{ij}\}$. We consider the affine $(N+2)$-space with coordinates x, y and $\{a_{ij}\}$, and we take $X \subset \mathbf{A}^{N+2}$ to be the hypersurface with equation $\sum a_{ij}x^iy^j = 0$. The projection map $\mathbf{A}^{N+2} \to \mathbf{A}^N = B$ restricts to a map $X \to B$, and it is easy to see that the fiber over a point of B is just the plane curve whose equation corresponds to that point.

Bearing in mind that a morphism of affine varieties corresponds to a homomorphism of rings in the opposite direction, we see that a family of algebras should be defined simply as an algebra: If S is an algebra over R, then for every maximal ideal $P \subset R$ we define the fiber over P to be the (R/P)-algebra S/PS. For arbitrary primes $P \subset R$ we define the fiber of S over P to be the $\kappa(P)$-algebra $\kappa(P) \otimes_R S$, where $\kappa(P) = K(R/P)$ denotes the quotient field of R/P as usual. This is a family of algebras parametrized by the maximal ideals of R. The algebra corresponding to the family of plane curves above is the algebra over $R := k[\{a_{ij}\}]$ defined by $S = R[x,y]/(\sum a_{ij}x^iy^j)$.

The problem with this definition is that it is too inclusive. The different fibers may have nothing to do with one another. Already in the preceding example of plane curves, the fiber over the point $0 \in B$ is the whole plane (the equation $f = \sum 0x^iy^j$ is identically 0)—not a curve at all! Our family must satisfy some conditions if it is to be worth studying.

In the geometry of manifolds, for example, one often restricts attention to families $\varphi : X \to B$ that are **locally trivial**, in the sense that for every point $x \in X$ there is a neighborhood U of x such that U is isomorphic to a product of $\varphi(U)$ and one of the fibers, $U \cong \varphi(U)$. In algebraic geometry and commutative algebra this idea still leads to an interesting definition, though to be most useful it must be applied with "neighborhoods" smaller than the Zariski neighborhoods introduced in Chapter 1 (the necessary ideas are introduced in Chapter 7). But many natural families do not fit into this framework.

If we look at the preceding example of plane curves, but exclude the fiber over 0, so that all the fibers are really curves, then it is reasonable to feel that the fibers have something to do with one another. Nevertheless, some fibers are singular curves while others are smooth. For instance consider the family of curves defined by equation $xy - a = 0$. As a varies, the family goes from a smooth hyperbola to a union of two lines (at $a = 0$). Such a family cannot be locally trivial near the point at which the two lines meet; but in many respects these curves all do belong together.

This suggests that there might be a more general notion of what a "good" family should be. The most inclusive in current use is that the family should be flat. In algebraic terms, where the family is represented by an R-algebra S, this means that S is flat over R (that is, flat as an R-module); In geometric terms, when the family is represented by a morphism $\varphi : X \to B$,

it means that for each point $x \in X$ there is an affine neighborhood U of X and an affine neighborhood V of $\varphi(x)$ such that φ restricts to a map $U \to V$, and the corresponding map of coordinate rings $A(V) \to A(U)$ makes $A(U)$ into a flat $A(V)$-module.

We now turn to some simple examples that give a feeling for flatness. Then we systematically investigate the algebraic properties of flatness. At the end of the chapter we explain the Rees algebra, a natural flat family that gives some insight into results such as Corollary 5.5, in which a good property of $\operatorname{gr}_I R$ was seen to "propagate" to R.

6.1 Elementary Examples

We take $R = k[t]$, the polynomial ring in one variable over an algebraically closed field, and look at some simple R-algebras to see which are flat.

Example 1 (Figure 6.1). $S = R[x]/(x-t)$. In this case $S \cong R$ is as good an R-algebra as one could possibly have. Since $R \otimes_R N = N$ for any R-module N, R is flat as an R-algebra.

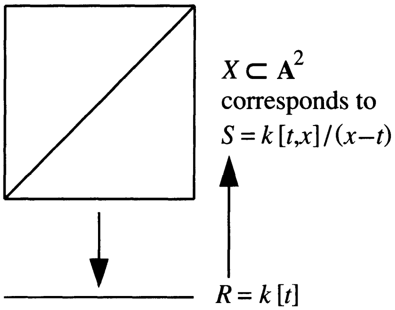


FIGURE 6.1.

Example 2 (Figure 6.2). $S = R[x]/(x^2 - t)$. In this case the fiber over a prime $P = (t-a)$, with $a \neq 0 \in k$, is

$$k[x]/(x^2 - a) \cong k \times k.$$

The fiber over (t) is $k[x]/(x^2)$. The fiber over (0) is $k(t)[x]/(x^2 - t)$, a field of degree 2 over the residue field $\kappa((0)) = k(t)$. We see that for each prime P the fiber over P is a vector space of dimension 2 over its residue field $\kappa(P)$. In fact S is a free R-module on the generators $(1, x)$, as the reader may check. Thus, $S \otimes_R N = N \oplus N$ for any R-module N, and it follows that S is flat.

Example 3 (Figure 6.3). $S = R[x]/(tx-1)$. We may identify S with $R[t^{-1}]$, so this example is obtained by localizing the algebra S in Example 1. As

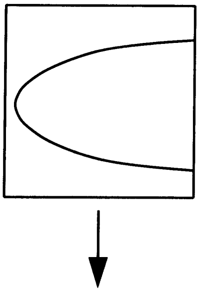


FIGURE 6.2.

we know from Proposition 2.5, localizations are flat, so this is again a flat family. The fiber over the prime $P \subset R$ is the $\kappa(P)$-algebra $\kappa(P)$, corresponding to one point, except when $P = (t)$, when the fiber is the zero ring, corresponding to the empty variety. Note that S is *not* a free R-module in this case.

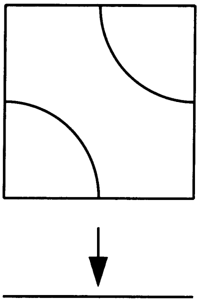


FIGURE 6.3.

Example 4 (Figure 6.4). $S = R[x]/(tx - t)$. In this example, S in *not* flat; we see that $t(x-1) = 0$ in S, so S has t-torsion, violating the criterion for flatness given in Corollary 6.3. It is also true that the fibers vary wildly: If the prime P does not contain t, then t is a unit in $\kappa(P)$ and thus also in $\kappa(P) \otimes_R S$, so

$$\begin{aligned} \kappa(P) \otimes_R S &= \kappa(P)[x]/(tx - t) \\ &= \kappa(P)[X]/(x - 1) \\ &\cong \kappa(P), \end{aligned}$$

but if $P = (t)$ then $tx - t = 0$ in $\kappa(P) \otimes_R R[x]$, so $\kappa(P) \otimes_R S = R[x]$, corresponding to the vertical line in Figure 6.4. (More generally, a morphism of varieties cannot be flat in the neighborhood of a point p if the fiber through p has dimension greater than that of nearby fibers; see Exercise 6.9 for an example, and Theorems 10.10, and 18.16 for more precise information.

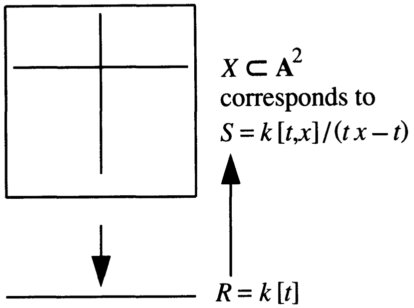


FIGURE 6.4.

For two further examples, see Exercises 6.8 and 6.9.

6.2 Introduction to Tor

We shall establish several criteria for the flatness of modules. Proposition 6.1 is the key to these criteria. For the statement and proof we shall make use, for the first time in this book, of the functors $\mathrm{Tor}_i^R(M,N)$ (actually only with $i = 0$ and 1). It is not hard to avoid the use of Tor in this proposition, and in this sense the proof can be made more elementary, but I feel that the elementary argument is more complicated and less transparent than the one with Tor. I have no doubt that those who are already familiar with Tor will agree; for the others, I think this is a good time to learn the elementary homological algebra required. It can be found, in a brief form, in Appendix 3. (For a more leisurely account the reader might consult Rotman [1979] or Hilton and Stammbach [1971].) So that the reader may judge the merits of the case for Tor, we present a proof of Proposition 6.1 without Tor as well.

For the purposes of this chapter, the reader needs to know the following about Tor:

1. If $\cdots \to F_{i+1} \to F_i \to F_{i-1} \to \cdots \to F_0 \to N \to 0$ is a free resolution of N as an R-module, then $\mathrm{Tor}_i^R(M,N)$ is the homology at $M \otimes F_i$ of the complex $M \otimes F_{i+1} \to M \otimes F_i \to M \otimes F_{i-1}$; that is, it is the kernel of $M \otimes F_i \to M \otimes F_{i-1}$ modulo the image of $M \otimes F_{i+1} \to M \otimes F_i$. This homology is independent of the resolution chosen. (We could also compute $\mathrm{Tor}_i^R(M,N)$ by tensoring N with a free resolution of M). From this it is very easy to deduce:

   a. $\mathrm{Tor}_0^R(M,N) = M \otimes_R N$.

b. If M or N is free, then $\operatorname{Tor}_i^R(M,N) = 0$ for $i > 0$ (the same is true for flat modules; see Exercise 6.1.).

c. As in the case of Tor_0, $\operatorname{Tor}_i^R(M,N)$ is a **covariant functor** of two R-modules M and N that is ***R*-bilinear** in the sense that it is an R-module, and the map "multiplication by $r \in R$" applied to either M or N induces "multiplication by r" on $\operatorname{Tor}_i^R(M,N)$.

d. If R is Noetherian and M and N are finitely generated R-modules, then $\operatorname{Tor}_i^R(M,N)$ is a finitely generated module.

e. If S is a flat R-algebra (such as a localization $R[U^{-1}]$), then $S \otimes_R \operatorname{Tor}_i^R(M,N) = \operatorname{Tor}_i^S(S \otimes_R M, S \otimes_R N)$.

2. For any short exact sequence $0 \to M' \to M \to M'' \to 0$ of R-modules, and any R-module N, there is a **long exact sequence of Tor**:

$$
\begin{array}{ll}
\cdots\cdots\cdots\cdots\cdots\cdots & \longrightarrow \\
\operatorname{Tor}_i^R(M',N) \longrightarrow \operatorname{Tor}_i^R(M,N) \longrightarrow \operatorname{Tor}_i^R(M'',N) & \longrightarrow \\
\cdots\cdots\cdots\cdots\cdots\cdots & \longrightarrow \\
\operatorname{Tor}_1^R(M',N) \longrightarrow \operatorname{Tor}_1^R(M,N) \longrightarrow \operatorname{Tor}_1^R(M'',N) & \longrightarrow \\
M' \otimes_R N \longrightarrow M \otimes_R N \longrightarrow M'' \otimes_R N \longrightarrow 0 &
\end{array}
$$

(With property 1b, this property actually characterizes Tor. See Exercise A3.15.)

Here is a useful example of a computation of Tor:

Suppose that $x \in R$ is a nonzerodivisor, and that M is an R-module. We shall compute the modules $\operatorname{Tor}_i^R(R/(x), M)$. The short exact sequence

$$0 \to R \xrightarrow{x} R \to R/(x) \to 0$$

is a actually a free resolution of $R/(x)$, and we use it in the definition of Tor. Thus the module $\operatorname{Tor}_i^R(R/(x), M)$ is the ith homology module of the complex

$$0 \to M \xrightarrow{x} M \to 0,$$

and we find

$$
\begin{aligned}
\operatorname{Tor}_0^R(R/(x), M) &= M/xM \\
\operatorname{Tor}_1^R(R/(x), M) &= (0 :_M x) \\
\operatorname{Tor}_i^R(R/(x), M) &= 0 \quad \text{for } i > 1.
\end{aligned}
$$

6.3 Criteria for Flatness

The relevance of Tor to flatness in general is exhibited in Exercise 6.1. For our purposes, a more specific result is more interesting.

Proposition 6.1. *Let R be a ring, and let M be an R-module. If I is an ideal of R, then the multiplication map $I \otimes_R M \to M$ is an injection iff* $\mathrm{Tor}_1^R(R/I, M) = 0$. *The module M is flat iff this condition is satisfied for every finitely generated ideal I.*

Proof. From the short exact sequence $0 \to I \to R \to R/I \to 0$, we get a long exact sequence containing

$$\mathrm{Tor}_1^R(R, M) \to \mathrm{Tor}_1^R(R/I, M) \to I \otimes M \to R \otimes M.$$

The left-hand term is 0 by property 1b, and the right-hand term is M. Additionally, the right-hand map is just the multiplication map $I \otimes M \to M$. The equivalence in the first assertion follows.

By definition, M is flat iff for every injection $N' \subset N$ of R-modules, the induced map $N' \otimes_R M \to N \otimes_R M$ is an injection. Suppose that the condition of the proposition (which is simply the special case where $N = R$ and $N' = I$ is finitely generated) is satisfied.

First we note that $I' \otimes M \to M$ is an injection for any ideal I' of R. Indeed, any element $0 \neq x$ of $I' \otimes_R M$ is a finite sum of elements $r' \otimes m$ for $r' \in I'$ and $m \in M$. Thus x comes from a necessarily nonzero element of some $I \otimes_R M$, with I finitely generated, so x goes to a nonzero element of M.

Similarly, the module $N \otimes_R M$ is generated by the elements $\{n \otimes_R m | n \in N, m \in M\}$, and the relations, which are the relations of bilinearity, each involve just finitely many elements of N. Thus, the statement that an $x \in N' \otimes_R M$ goes to 0 in $N \otimes_R M$ involves only finitely many elements of N, and we may assume that N is finitely generated.

Now we can find a sequence of submodules

$$N' = N_0 \subset N_1 \subset \cdots \subset N_p = N$$

such that each N_{i+1}/N_i is generated by one element. Of course, it suffices to show that each $N_i \otimes M \to N_{i+1} \otimes M$ is injective, so we may assume from the outset that N/N' is a cyclic module, and write $N/N' \cong R/I$.

The short exact sequence $0 \to N' \to N \to N/N' \to 0$ gives rise to a long exact sequence containing the terms

$$\mathrm{Tor}_1^R(N/N', M) \to N' \otimes M \to N \otimes M.$$

Since $\mathrm{Tor}_1^R(N/N', M) = \mathrm{Tor}_1^R(R/I, M) = 0$ by hypothesis, we are done. □

We used Tor in this proof to show that if $I \subset R$ is an ideal and $I \otimes M \to R \otimes M = M$ is an inclusion, then $N' \otimes M \to N \otimes M$ is an inclusion for any

modules $N' \subset N$ such that $N/N' \cong R/I$. For purposes of comparison, we shall now prove just this point without using Tor.

Choose an element $n \in N$ that maps to $1 \bmod I$ in $N/N' \cong R/I$, and let $\varphi : R \to N$ be the map sending 1 to n. The map φ carries I into N'; write $\varphi' : I \to N'$ for the induced map. The kernel of φ is contained in I. Since the induced map of R/I to N/N' is an isomorphism, the cokernels of φ and φ' coincide. Writing C for the cokernel, we thus get a commutative diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
0 & \to & I & \to & R & \to & R/I & \to & 0 \\
 & & {\scriptstyle\varphi'}\downarrow & & \downarrow{\scriptstyle\varphi} & & \| & & \\
0 & \to & N' & \to & N & \to & R/I & \to & 0 \\
 & & \downarrow & & \downarrow & & & & \\
 & & C & = & C & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

Tensoring with M, the upper row remains exact by hypothesis, and the columns remain exact by the right exactness of the tensor product. It follows that the kernels of $\varphi' \otimes M$ and $\varphi \otimes M$ coincide—call them K. We thus have a diagram with exact columns:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & K & = & K & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & I \otimes M & \to & R \otimes M & \to & (R/I) \otimes M & \to & 0 \\
 & & {\scriptstyle\varphi'\otimes 1}\downarrow & & \downarrow{\scriptstyle\varphi\otimes 1} & & \| & & \\
 & & N' \otimes M & \to & N \otimes M & \to & (R/I) \otimes M & \to & 0 \\
 & & \downarrow & & \downarrow & & & & \\
 & & C \otimes M & = & C \otimes M & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

If an element $x \in N' \otimes M$ goes to 0 in $N \otimes M$, then it also goes to 0 in $C \otimes M$, and thus comes from an element $y \in I \otimes M$. This element, regarded as an element of $R \otimes M$, goes to 0 under $\varphi \otimes 1$, and thus comes from K. Thus $x = (\varphi' \otimes 1)(y) = 0$. This shows that the map from $N' \otimes M$ to $N \otimes M$ is a monomorphism as required.

If the reader compares this last argument with the last paragraph of the proof of Proposition 6.1, it will be clear at least that a considerable compression has occurred even in this very simple case. I feel that there is also an increase in clarity.

It seems natural to ask whether one could refine Proposition 6.1 by limiting the ideals I that must be tested. In the most common situation it is enough to check for all maximal ideals (Theorem 6.8).

Before deriving the general characterization of flatness that comes out of Proposition 6.1, we give the simplest—and most useful—special cases.

Corollary 6.2. *Let k be a field. If $R = k[t]/(t^2)$, and M is an R-module, then M is flat iff multiplication by t from M to tM induces an isomorphism $M/tM \to tM$.*

Proof. The only nontrivial ideal of R is (t), which is isomorphic as an R-module to $R/(t)$ by the map $R/(t) \to (t)$ sending 1 to t. Applying the criterion of Proposition 6.1, we see that M is flat iff the map

$$M/tM \cong R/(t) \otimes_R M \cong (t) \otimes_R M \to R \otimes_R M = M,$$

sending the class in M/tM of an element $m \in M$ to $tm \in M$, is a monomorphism. But this map is the composition of an epimorphism $M/tM \to tM$ induced from multiplication by t with the inclusion $tM \subset M$. □

Corollary 6.3. *If $a \in R$ is a nonzerodivisor in R, and M is a flat R-module, then a is a nonzerodivisor on M. If R is a principal ideal domain, then the converse is also true: M is flat as an R-module iff M is torsion free.*

Proof. Let $I = Ra$ with $a \in R$ a nonzerodivisor. If M is flat, then for any $I \subset R$ the map $\varphi : I \otimes_R M \to R \otimes_R M = M$ is an injection. But $I \cong R$ by an isomorphism sending a to 1. Identifying $I \otimes_R M$ with $R \otimes_R M = M$ by means of this isomorphism, the map φ becomes multiplication by a. By definition, this is a monomorphism iff a is a nonzerodivisor on M.

If now R is a principal ideal domain, then every nonzero ideal is generated by a nonzerodivisor. The condition of Proposition 6.1 is trivial in the case $I = 0$, so we see from the computation of $\mathrm{Tor}_i^R(R/x, M)$ given at the end of section 6.2 that M is flat iff every nonzero element of R is a nonzerodivisor on M. This is what it means for M to be torsion-free. □

Before tackling the general case, we need an equational characterization of when an element of a tensor product of modules is zero. This is a translation of the fact that tensor products preserve direct sums and right-exact sequences.

Lemma 6.4. *Let M and N be R-modules, and suppose that N is generated by a family of elements $\{n_i\}$. Every element of $M \otimes_R N$ may be written as a finite sum $\sum_i m_i \otimes n_i \in M \otimes_R N$. Such an expression is 0 iff there exist elements m'_j of M and elements a_{ij} of R such that*

$$\sum_j a_{ij} m'_j = m_i \quad \textit{for all } i$$

and

$$\sum_i a_{ij} n_i = 0 \text{ in } N \quad \text{for all } j.$$

Proof. If elements m'_j and a_{ij} with the specified properties exist, then $\sum_i m_i \otimes n_i = \sum_i (\sum_j a_{ij} m'_j) \otimes n_i = \sum_j (m'_j \otimes \sum_i a_{ij} n_i) = 0$.

To prove the converse, we begin with what amounts to a special case: If N is free and the n_i are a basis, then $\sum_i m_i \otimes n_i = 0$ iff all the m_i are 0. This follows because $N = \oplus R n_i$, so

$$M \otimes_R N \cong \oplus_i (M \otimes_R R n_i) \cong \oplus_i M,$$

and $\sum_i m_i \otimes n_i$ corresponds to the vector $(m_1, m_2, \ldots)$.

Passing to the general case, let $F \to G \to N \to 0$ be a free presentation of N, chosen so that a basis $\{g_i\}$ of G maps to the set of elements $\{n_i\}$, say $g_i \mapsto n_i$. By the right-exactness of the tensor product functor, the natural sequence

$$M \otimes_R F \to M \otimes_R G \to M \otimes_R N \to 0$$

is exact, and of course $\sum_i m_i \otimes g_i$ goes to zero. It follows that $\sum_i m_i \otimes g_i = \sum_j m'_j \otimes y_j$ for some $m'_j \in M$, with y_j in the image of F, that is, with $y_j \to 0$ in N. We may write each y_j in terms of the basis g_i, say $y_j = \sum_{i=1}^{r} a_{ij} g_i$. But using the special case above on the difference $0 = \sum_i m_i \otimes g_i - \sum_j m'_j \otimes (\sum_i a_{ij} g_i)$, we see that $m_i = \sum_j a_{ij} m'_j$, while $y_j = \sum_i a_{ij} g_i$ goes to $0 = \sum_i a_{ij} n_i$, as required. □

It is crucial for the truth of each half of this lemma that the n_i actually generate N. For example, consider a contrary case, where $R = k[t]$ and $N = k[t]/(t^2)$, with $n_1 = t$. Taking $M = k[t]/(t)$, and $m_1 = 1$, we see that the element $m_1 \otimes n_1 = 1 \otimes t \in k[t]/(t) \otimes_{k[t]} k[t]/(t^2)$ is 0, but m_1 cannot be expressed as $a m'_1$ in such a way that $a n_1 = 0$. Of course, the general reason is that if the same criterion held for a set of elements n_i that generate only a submodule of N, then the inclusion map of this submodule into N would remain a monomorphism when tensored with M! This remark can be put to work to derive a characterization of flatness by equations that generalizes the preceding cases.

Corollary 6.5 (Equational Criterion for Flatness). *An R-module M is flat iff the following condition is satisfied:*

For every relation $0 = \sum_i n_i m_i$ with $m_i \in M$ and $n_i \in R$ there exist elements $m'_j \in M$ and elements $a_{ij} \in R$ such that

$$\sum_j a_{ij} m'_j = m_i \quad \text{for all } i$$

and

$$\sum_i a_{ij} n_i = 0 \text{ in } R \quad \text{for all } j.$$

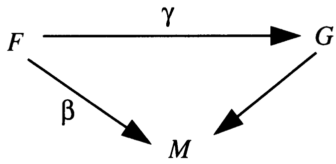
Proof. The condition of Proposition 6.1 may be restated as saying that M is flat if for every ideal I, an element $x = \sum_i n_i \otimes m_i \in I \otimes_R M$ goes to 0 in $R \otimes_R M$ iff $\sum_i n_i \otimes m_i$ satisfies the criterion of Lemma 6.4 for being 0. The image of x in $R \otimes_R M = M$ is just $\sum_i n_i m_i$, so the desired result follows.$\square$

Finally, the characterization of flatness can be reformulated in terms of maps of modules in an appealing way. In the language of Exercise 4.11 this even shows that finitely presented flat modules are the same as finitely generated projective modules.

Corollary 6.6. *Let R be a ring, and let M be an R-module. The following conditions are equivalent:*

a. M is flat.

b. For every map $\beta : F \to M$ from a finitely generated free module F, and for every submodule K of $\ker \beta$ *generated by one element, there is a commutative diagram*



with G free such that $K \subset$ $\ker \gamma$.

c. The same as statement b, but for arbitrary finitely generated submodules K of $\ker \beta$.

In particular, if M is finitely presented, then M is flat iff M is a summand of a free module.

Proof. a$\Longleftrightarrow$b: This is a "diagrammatic" translation of Corollary 6.5. An element f in the kernel of a map from a free module F to M is a relation on the images $m_j \in M$ of the basis elements of F. The elements m'_j of Corollary 6.5 correspond to a map from another free module, G, taking the generators of G to the m'_j. A matrix with entries a_{ij} such that $\sum_j a_{ij} m'_j = m_j$ corresponds to a map γ, making the diagram commute. The condition that $\sum_i a_{ij} n_i = 0$ in R for all j says that $\gamma(f) = 0$.

b$\Longleftrightarrow$c: Suppose we can find a map γ whose kernel contains a given element of K. Composing γ with a map killing the image of another element of K, and continuing in this way, we finally arrive at a map whose kernel contains K. The other implication is trivial.

For the last statement, note first that direct sums and direct summands of flat modules are flat. Since R itself is flat (tensoring with R is the identity functor), any summand of a free module is flat.

To say that M is finitely presented means that there is a surjection $F \to M$ from a finitely generated free module having kernel K finitely generated. If M is flat, let γ be as in statement b. The image of γ is carried isomorphically to M by the map from G. Thus the map $G \to M$ splits, so M is a direct summand of G. □

See Exercise 6.2 for another proof of the last statement of this corollary. We shall extend this criterion to the statement that a module is flat iff it is the "filtered direct limit" of free modules (Govorov-Lazard theorem) in Appendix 6.

If (R, P) is a local ring and S is a flat R algebra, then good properties of the fiber S/PS over R/P often imply good properties of S. We shall see some dimension-theoretic versions of this statement in Chapters 10 and 14. For now we prove a result that is a generalization of Corollary 5.5.

Corollary 6.7. *Let k be a field, let $R = k[t]$ be the polynomial ring in one variable, and let S be a Noetherian ring that is flat over R. If the fiber S/tS over the prime (t) is a domain, and U is the set of elements of the form $1 - ts$ for $s \in S$, then $S[U^{-1}]$ is a domain.*

Proof. We may replace S by $S[U^{-1}]$ at the outset and assume that all elements of the form $1 + st$ are units of S.

Suppose that $I, J \subset S$ are ideals with $IJ = 0$; we must show that either I or J is 0. Enlarging I and J if necessary, we may assume that each is the annihilator of the other. Since $IJ \equiv 0(\text{modulo } t)$, and $S/(t)$ is a domain, we may suppose $J \subset (t)$. Thus $J = (J : t)t$. Since t is a nonzerodivisor and $I(J : t)t = 0$, it follows that $(J : t)$ annihilates I. Thus $(J : t) \subset J$, and we have $J = Jt$. By Corollary 4.7 there is an element $s \in S$ such that $(1 - ts)J = 0$, so $J = 0$. □

It follows from primary decomposition that in the setting of Corollary 6.7 we can find a single element $f \in U$ such that $S[f^{-1}]$ is a domain. However, one may not be able to avoid localization completely, as one sees from the example $R = k[t] \to S = k[x, t] \times k[t, t^{-1}]$, where k is a field, pictured in Figure 6.5. The fiber over the maximal ideal $(t-a)$ (for $a \in k$) is $S/(t-a)S$. When $a = 0$ this fiber is a domain since $tk[t, t^{-1}] = k[t, t^{-1}]$, but for all a other than 0 the fiber is not a domain. Such troubles can be avoided by working with graded rings—geometrically, working with projective maps, or, more generally, with proper maps.

6.4 The Local Criterion for Flatness

We have shown in Proposition 6.1 that an R-module M is flat iff the maps $I \otimes M \to M$ are injections for all ideals I of R. We shall now show that if

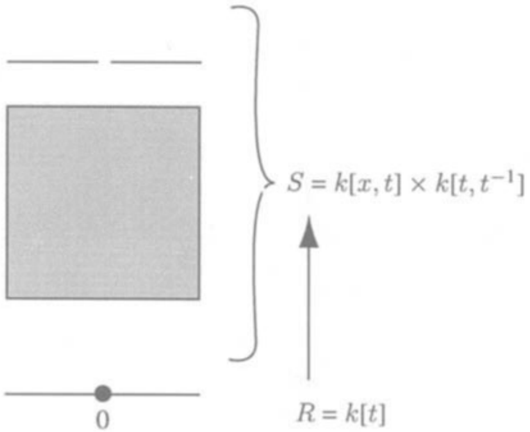


FIGURE 6.5.

R is a local ring and M satisfies some mild hypotheses, then it suffices to check this condition when I is the maximal ideal of R.

Theorem 6.8 (Local Criterion for Flatness). *Suppose that $(R, \mathfrak{m})$ is a local Noetherian ring, and let $(S, \mathfrak{n})$ be a local Noetherian R-algebra such that $\mathfrak{m}S \subset \mathfrak{n}$. If M is a finitely generated S-module, then M is flat as an R-module iff* $\operatorname{Tor}_1^R(R/\mathfrak{m}, M) = 0$.

The theorem is often applied with $M = S$, to establish the flatness of S. See Theorem 18.16 and Exercise 18.17 for typical cases. However, it is also interesting for the case $R = S$ to test the flatness of a finitely generated module. In this case the result is both easier and sharper; see Exercise 6.2. For the necessity of the hypothesis, see Exercise 6.3. For a different-looking statement proved by almost the same argument, see Exercise 6.5.

Proof. If M is flat then $\operatorname{Tor}_1^R(R/\mathfrak{m}, M) = 0$, by Proposition 6.1.

Now suppose that S and M are as in the theorem, and that $\operatorname{Tor}_1^R(R/\mathfrak{m}, M) = 0$. As a preliminary step we shall show that if N is an R-module of finite length, then $\operatorname{Tor}_1^R(N, M) = 0$. We may prove this by induction on the length, the hypothesis being the case of length 1: If N' is any proper submodule of N, then the exact sequence $0 \to N' \to N \to N/N' \to 0$ gives rise to an exact sequence of Tor containing the terms

$$\operatorname{Tor}_1^R(N', M) \to \operatorname{Tor}_1^R(N, M) \to \operatorname{Tor}_1^R(N/N', M).$$

By induction on the length, $\operatorname{Tor}_1^R(N', M) = 0 = \operatorname{Tor}_1^R(N/N', M)$, so $\operatorname{Tor}_1^R(N, M) = 0$ as required.

Now let I be an arbitrary ideal, and suppose that $u \in I \otimes_R M$ is in the kernel of the multiplication map $I \otimes_R M \to M$. We shall prove that $u = 0$.

The S-module structure on M gives $I \otimes_R M$ the structure of an S-module too, and we have $\mathfrak{m}^n(I \otimes_R M) \subset \mathfrak{n}^n(I \otimes_R M)$. It is finitely generated as an S-module, so by the Krull intersection theorem (Theorem 5.4), $\cap_n \mathfrak{n}^n(I \otimes_R M) = 0$, and we see that $\cap_n \mathfrak{m}^n(I \otimes_R M) = 0$. Thus it suffices to show that $u \in \mathfrak{m}^n(I \otimes_R M)$ for every n (this is the only use we shall make of the hypothesis on M).

The module $\mathfrak{m}^n(I \otimes_R M)$ is the image in $I \otimes_R M$ of $(\mathfrak{m}^n I) \otimes_R M$. By the Artin-Rees lemma, $\mathfrak{m}^t \cap I \subset \mathfrak{m}^n I$ for sufficiently large t, so it suffices to show that u is in the image of $(\mathfrak{m}^t \cap I) \otimes_R M$ for all t. Tensoring the short exact sequence

$$0 \to \mathfrak{m}^t \cap I \to I \to I/(\mathfrak{m}^t \cap I) \to 0$$

with M produces the exact sequence

$$(\mathfrak{m}^t \cap I) \otimes_R M \to I \otimes_R M \to I/(\mathfrak{m}^t \cap I) \otimes_R M \to 0.$$

It thus suffices to show that u goes to 0 in $I/(\mathfrak{m}^t \cap I) \otimes_R M$. The map $I \otimes_R M \to I/(\mathfrak{m}^t \cap I) \otimes_R M$ is obtained by tensoring the top row of the commutative diagram

$$\begin{array}{ccc} I & \to & I/(\mathfrak{m}^t \cap I) \\ \downarrow & & \downarrow \varphi \\ R & \to & R/\mathfrak{m}^t \end{array}$$

with M to get

$$\begin{array}{ccc} I \otimes_R M & \to & I/(\mathfrak{m}^t \cap I) \otimes_R M \\ \downarrow & & \downarrow \varphi\otimes 1 \\ M = R \otimes_R M & \to & R/\mathfrak{m}^t \otimes_R M. \end{array}$$

Since u goes to zero under the left-hand vertical map, we see that it suffices to show that the kernel of the right-hand vertical map $\varphi \otimes 1$ is 0.

Identifying $I/(\mathfrak{m}^t \cap I)$ with $(I + \mathfrak{m}^t)/\mathfrak{m}^t$, we see that φ is the left-hand map in the short exact sequence

$$0 \to (I + \mathfrak{m}^t)/\mathfrak{m}^t \to R/\mathfrak{m}^t \to R/(I + \mathfrak{m}^t) \to 0.$$

Applying Tor, we get a long exact sequence of which a part is

$$\mathrm{Tor}_1^R(R/(I + \mathfrak{m}^t), M) \to (I + \mathfrak{m}^t)/\mathfrak{m}^t \otimes_R M \xrightarrow{\varphi\otimes 1} R/\mathfrak{m}^t \otimes_R M,$$

so it is enough to show that $\mathrm{Tor}_1^R(R/(I + \mathfrak{m}^t), M) = 0$. Since $R/(I + \mathfrak{m}^t)$ is annihilated by $\mathfrak{m}^t$, it is a module of finite length, and we are done. $\square$

If $R \to R'$ is any homomorphism of rings and M is a flat R-module, then $R' \otimes_R M$ is flat as an R'-module because tensoring over R' with $(R' \otimes_R M)$ is the same as tensoring over R with M—that is,

$$(R' \otimes_R M) \otimes_{R'} N = M \otimes_R N$$

for any R'-module N. We can use the criterion to prove the converse in an important special case:

Corollary 6.9. *Suppose that $(R, \mathfrak{m})$ is a local Noetherian ring. Let $(S, \mathfrak{n})$ be a local Noetherian R-algebra such that $\mathfrak{m}S \subset \mathfrak{n}$, let $x \in \mathfrak{m}$ be a nonzerodivisor on R, and let M be a finitely generated S-module. If x is a nonzerodivisor on M, then M is flat over R iff M/xM is flat over $R/(x)$.*

Proof. If M is flat then $M/xM = R/(x) \otimes_R M$ is flat over $R/(x)$ without any hypothesis, as we have just remarked, so we suppose that M/xM is flat over $R/(x)$, and prove that M is flat over R. Let $k = R/\mathfrak{m}$ be the residue class field of R.

We have $\operatorname{Tor}_1^{R/(x)}(k, M/xM) = 0$ since M/xM is flat over $R/(x)$. By Lemma 6.10, $\operatorname{Tor}_1^R(k, M) = 0$, so M is flat by the local criterion, Theorem 6.8. □

Lemma 6.10. *If R is a ring, M is an R-module, and $x \in R$ is a nonzerodivisor on R and on M, then for any $(R/(x))$-module N we have $\operatorname{Tor}_i^{R/(x)}(N, M/xM) = \operatorname{Tor}_i^R(N, M)$.*

Proof. Let

$$\mathfrak{F}: \quad \cdots \to F_2 \to F_1 \to F_0$$

be a free resolution of M as an R-module. We claim that $R/(x) \otimes \mathfrak{F}$ is a free resolution of $R/(x) \otimes M$. Given this, we may compute $\operatorname{Tor}_i^{R/(x)}(N, M/xM)$ as the homology of $N \otimes_{R/(x)} R/(x) \otimes_R \mathfrak{F} \cong N \otimes_R \mathfrak{F}$, and we see that it coincides with $\operatorname{Tor}_i^R(N, M)$, as required.

The homology of the complex

$$R/(x) \otimes \mathfrak{F}: \quad \cdots \to F_2/xF_2 \to F_1/xF_1 \to F_0/xF_0$$

is $\operatorname{Tor}^R(R/(x), M)$. As we have shown at the end of section 6.2, we have $\operatorname{Tor}_0^R(R/(x), M) = M/xM$, and since x is a nonzerodivisor on R, $\operatorname{Tor}_i^R(R/(x), M) = 0$ for all $i > 0$. This is the same as saying that the complex $R/(x) \otimes \mathfrak{F}$ is a resolution of M/xM. Since the modules F_i/xF_i are free over $R/(x)$, we are done. □

The corollary is often used in the following situation: Suppose that

$$X \xrightarrow{\varphi} Y \xrightarrow{\psi} \mathbf{A}^1$$

are maps of affine varieties over a field k, and that X and Y are flat over $\mathbf{A}^1$. We say that φ is flat, or that X is flat over Y, if the corresponding map from the coordinate ring $A(Y)$ of Y to the coordinate ring $A(X)$ of X makes $A(X)$ into a flat $(A(Y))$-module. For each point $p \in \mathbf{A}^1$ we have

a map of "fibers" $X_p := (\psi\varphi)^{-1}(p) \to \psi^{-1}(p) =: Y_p$, as in Figure 6.6 where we have chosen points $p' \in Y$ mapping to p and $p'' \in X$ mapping to p'. We think of the whole setup as a family of maps of varieties.

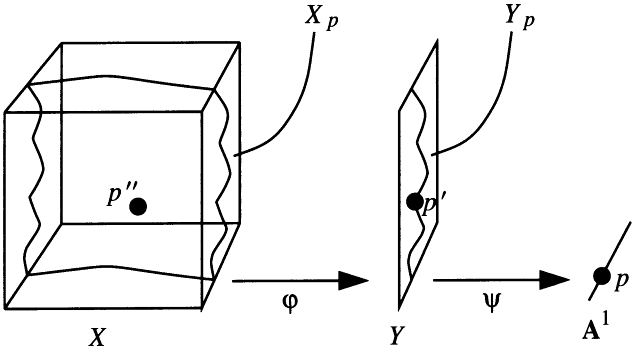


FIGURE 6.6.

Let x be a coordinate function on $\mathbf{A}^1$ that takes the value 0 at p, and let $R = A(Y)_{P'}$, $M = A(X)_{P''}$, where P' and P'' are the maximal ideals corresponding to p' and p'', respectively. If we assume that (as in the picture) X and Y map onto an open set of $\mathbf{A}^1$, then the maps $k[x] \to R$ and $k[x] \to M$ corresponding to ψ and to $\psi\varphi$, respectively, are both injections from $k[x]$ into the domains R and M, so that x is a nonzerodivisor on both R and M. By Corollary 6.3, X and Y are flat over $\mathbf{A}^1$.

In this setting Corollary 6.9 may be interpreted as saying that if the map $X_p \to Y_p$ is flat in a neighborhood of p'' in X_p, then $X \to Y$ is flat in a neighborhood of p'' in X. That is, if the fibers of $X_p \to Y_p$ vary "nicely" near p'', then the same is true of all the fibers of $X \to Y$ near p''.

6.5 The Rees Algebra

One way of producing flat families is through the technique of Rees algebras. Let R be a ring and let $I \subset R$ be an ideal. We define the **Rees algebra of R with respect to I** to be the R-algebra

$$\mathcal{R}(R, I) := \sum_{n=-\infty}^{\infty} I^n t^{-n} = R[t, t^{-1}I] \subset R[t, t^{-1}],$$

where we take $I^n = R$ if $n \leq 0$. If R is a k-algebra, then we regard $\mathcal{R}(R, I)$ as a $k[t]$-algebra. It is then clear that

$$\mathcal{R}(R, I)/t\mathcal{R}(R, I) = \operatorname{gr}_I R$$
$$\mathcal{R}(R, I)/(t - a)\mathcal{R}(R, I) = R \quad \text{for any } 0 \neq a \in k;$$

that is, the Rees algebra defines a family of k-algebras over the line with parameter t having fiber gr_I R at $t = 0$ and fiber R at $t = a$ for every $a \neq 0$. We shall show that the Rees algebra is flat over $k[t]$. Combined with the second statement of the following result and Corollary 6.7, this gives an alternative proof of Corollary 5.5, much longer than the original to be sure, but suggesting a general technique.

Corollary 6.11. *If R is an algebra over a field k, then the Rees algebra $S = \mathcal{R}(R, I)$ is flat over $k[t]$. If $\cap_{d=1}^{\infty} I^d = 0$, then every element of the form $1 - ts$ with $s \in S$ is a nonzerodivisor on $\mathcal{R}(R, I)$.*

Proof. For the first statement it is enough by Corollary 6.3 to show that $\mathcal{R}(R, I)$ is torsion-free as a $k[t]$-module. Since $\mathcal{R}(R, I) \subset R[t, t^{-1}]$, this is immediate. For the second statement, note that if $p(1 - ts) = 0$ for some s in S, then, reading the equation modulo t, we must have $p = qt$ for some $q \in S$. But t is a nonzerodivisor on S, so $q(1 - ts) = 0$. Repeating this argument, it follows that $p \in t^n S$ for every n. Writing $p = \sum_{i=-j}^{j} p_i t^i$, with $p_i \in R$, we see that $p_i \in I^n$ for each n, so $p = 0$ as required. □

6.6 Exercises

Exercise 6.1 (Tor and flatness):* Here is the basic relation between Tor and flatness:

a. Let R be a ring, and let M be an R-module. Show that M is flat iff $\mathrm{Tor}_1^R(M, N) = 0$ for all R-modules N iff $\mathrm{Tor}_1^R(N, M) = 0$ for all R-modules N.

b.* Show that M is flat iff $\mathrm{Tor}_i^R(M, N) = 0$ for all R-modules N and all $i > 0$.

Exercise 6.2 (Finitely presented flat modules are locally free):* Let R be a ring and let M be a finitely presented R-module. Show that the following statements are equivalent:

i. M is flat over R.

ii. M_P is flat over R_P for all maximal ideals P of R.

iii. $\mathrm{Tor}_1^R(M, R/P) = 0$ for all maximal ideals P of R.

iv. M_P is a free module over R_P for all maximal ideals P of R.

v. M is a projective R-module.

Here are some steps that may help you:

a.* Let R, P be a local ring, and use Nakayama's lemma to show that if M is a finitely presented module, then M is flat iff M is free iff $\operatorname{Tor}_1^R(M, R/P) = 0$.

b. Prove the equivalence of statements i and ii by localizing; one way to do it is to show that if N is an R_P-module then $M \otimes_R N = M \otimes_{R_P} N$, and use Corollary 2.9; another is to show that Tor localizes.

c. Use parts a and b with Exercise 4.11 to show that statements i, iv, and v are equivalent. Statement i implies ii by Exercise 6.1, and the converse comes by localizing and using part a.

Exercise 6.3:* Let $R = k[x]_{(x)}$, where k is a field, and let M be any R-module. Show that $\operatorname{Tor}_1^R(R/(x), M) = 0$ iff M is flat, so that Theorem 6.8 holds for this ring without restriction on M. Show, however, that if $R = k[x, y]_{(x,y)}$ and $M = k(x)$ (with y acting as 0), then $\operatorname{Tor}_1^R(R/(x, y), M) = 0$ but M is not flat.

Exercise 6.4:* Let $S = R[x_1, \ldots, x_r]$ be the polynomial ring in r variables over a Noetherian ring R, and let $f \in S$ be a nonzerodivisor (see Exercise 3.4). Show that $S/(f)$ is a flat R-module iff the coefficients of f generate the unit ideal of R. In case $R = k[x]$, $S = R[y]$, and $f = 1 + xy$, show that $S/(f)$ is not free as an R-module.

Exercise 6.5 (Infinitesimal criterion of flatness): Prove the following by adapting the proof of Theorem 6.8: Suppose that (R, P) is a local Noetherian ring, and let (S, Q) be a local Noetherian R-algebra such that $PS \subset Q$. If M is a finitely generated S-module, show that M is flat as an R-module iff M/P^nM is flat as an R/P^n-module for every n.

Exercise 6.6 (The family of projective plane curves): The algebra in this exercise corresponds to the first flat family ever considered (implicitly, of course; to translate this exercise directly into geometry requires an algebraically closed field, so the time we are speaking of is about 1830, when people were first seriously investigating projective plane curves over $\mathbf{C}$). It is still an object of active research. Fix a degree d. For each 3-component, multiindex $\alpha = (a_0, a_1, a_2)$ of degree d (that is, the a_i are nonnegative integers with $a_0 + a_1 + a_2 = d$) let x_α be an indeterminate. Let $R = k[\{x_\alpha\}]$ be the polynomial ring in the x_α, and let S be the R-algebra

$$R[y_0, y_1, y_2]/\sum_\alpha x_\alpha y^\alpha,$$

where we have written y^α for the monomial $y_0^{a_0} y_1^{a_1} y_2^{a_2}$. Geometrically, this corresponds to the family of all projective plane curves of degree d; of course one could replace 3 by any number $r+1$ and get the family of hypersurfaces

of degree d in $\mathbf{P}^r$. Except for the fiber over the point where all the x_α are zero, this family is "good": The geometric properties of two plane curves of degree d are closely related, and algebraically the fiber at a prime P is a polynomial ring over $\kappa(P)$ modulo an equation of degree d; certainly these have a strong "family resemblance." Show that if we invert any x_α, the family becomes flat, that is, the family given by the $R[x_\alpha^{-1}]$-algebra $S[x_\alpha^{-1}]$ is flat, by showing that it is a free R-module (not finitely generated!). Show that S is an integral domain (that is, $\sum_\alpha x_\alpha y^\alpha$ is prime) and contains R, so that S is torsion-free as an R-module. Show, however, that S itself is not flat over R by proving and using the following facts:

a.* If S is a flat module over a ring R, and $R \to T$ is any map of rings, then $S \otimes_R T$ is flat over T.

b. There is a map of rings $R = k[\{x_\alpha\}] \to k[t] = T$ such that $T \otimes_R S = k[t, y_0, y_1, y_2]/ty_0^d$, and this is not a flat T-algebra.

Exercise 6.7 (Flatness and (almost) regular sequences): If R is not a principal ideal domain, then the condition of Corollary 6.3 is not sufficient for flatness, as the example of Exercise 6.6 shows. However, the idea of part a of that exercise proves a much more powerful consequence that is the first hint of the important interaction of flatness and regular sequences. Show that if S is a flat R-module, and $x_1, \ldots, x_r$ is a sequence of elements of R such that for each i the element x_i is a nonzerodivisor on $R/(x_1, \ldots, x_{i-1})$, then for each i the element x_i is a nonzerodivisor on $S/(x_1, \ldots, x_{i-1})$. (This is not quite the condition that $x_1, \ldots, x_r$ be a regular sequence (see Chapter 10), because we are not insisting that $(x_1, \ldots, x_r)R \neq R$ or that $(x_1, \ldots, x_r)S \neq S$.) Although we are not ready to prove it, this condition is actually equivalent to flatness in many cases of interest, for example in the case where R is a polynomial ring and S is local, with maximal ideal containing $x_1, \ldots, x_r$ (see Exercise 18.18).

Exercise 6.8: Let k be a field, and set $R = k[t]$, $S = R[x]/((x) \cap (x,t)^2)$ should be $((x) \cap (x,t)^2)$ (see Figure 6.7). Show that in this example, S is not flat over R. (You may use the criterion for flatness given in Corollary 6.3.)

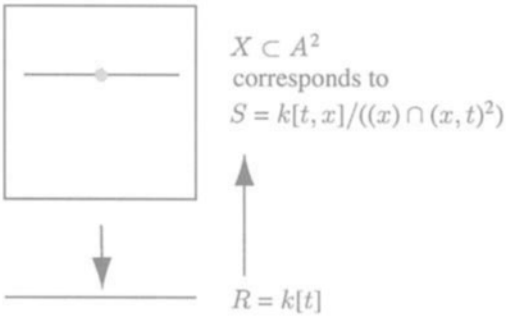


FIGURE 6.7.

Exercise 6.9: Use Exercise 6.7 to show that the blowup of the plane is not flat over the plane: That is, if $R = k[x,y]$ and S is the subring of the quotient field of R generated by the two elements x/y and y (crudely, $S = k[x/y, y]$), then S is not flat as an R-module. The inclusion $R \subset S$ corresponds to the map from the plane to the plane suggested by Figure 6.8.

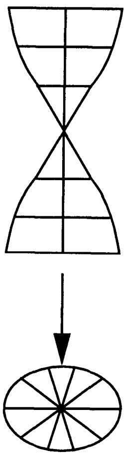


FIGURE 6.8. Blowup of a point in the plane.

Intuitively, flatness fails because the fiber over the origin is a curve, whereas nearby fibers are only points (see Theorem 10.10 and Exercise 10.5 for a more precise treatment).

Exercise 6.10 (Flatness of graded modules):* Let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded ring with R_0 a field, and let M be a graded R-module.

a. Show that M is flat over R iff $I \otimes M \to M$ is an injection for every homogeneous ideal I of R.

b. Show that M is flat iff M_P is flat, where $P = R_1 \oplus R_2 \oplus \cdots$ is the homogeneous maximal ideal.

Flat Families of Graded Modules

Exercise 6.11: Let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded ring such that R_0 is a local Noetherian ring and that R is finitely generated as an R_0-algebra by elements of degree 1. Let M be a finitely generated graded R-module, and let M_d be the degree d part of M. We may think of M as a family of graded modules over the "base" R_0. If $R = R_0$, then we have seen that M is flat over R_0 iff M is free over R_0. The general case, which arises often, can be analyzed in terms of this one:

a. Show that each M_d is finitely generated as an R_0-module.

b. Show that M is flat over R_0 iff M_d is free over R_0 for all d.

c.* (The following exercise may be interpreted as describing flatness for families of sheaves on projective space. See Hartshome [1977], Chapter II, for more information.) For every $f \in R_1$ the localization $M[f^{-1}]$ is graded; we write $M[f^{-1}]_0$ for its component of degree 0. Show that $M[f^{-1}]_0$ is flat over R_0 for all $f \in R_1$ iff M_d is a free R_0-module for all $d \gg 0$.

d. For each prime P of R_0 we may define a Hilbert function

$$H_{\kappa(P)\otimes M}(d) = \dim_{\kappa(P)} \kappa(P) \otimes M_d.$$

As in Chapter 1 we write $P_{\kappa(P)\otimes M}(d)$ for the polynomial in d that agrees with $H_{\kappa(P)\otimes M}(d)$ for large d (see also Chapter 12). Show that if M is flat over R_0, then the function $H_{\kappa(P)\otimes M}(d)$ is independent of the prime P chosen in R_0. Show that if $M[f^{-1}]_0$ is flat over R_0 for all $f \in R_1$, then the polynomial $P_{\kappa(P)\otimes M}(d)$ is independent of P.

As we shall see in Chapter 12, the Krull dimension (see Chapter 9 for the definition) of $\kappa(P)\otimes M$ as a module over $\kappa(P)\otimes R$ can be read off from this polynomial, so a consequence of flatness is that the Krull dimension of $\kappa(P) \otimes M$ is constant. We shall further see, in Exercise 20.14, that in this situation the constancy of the Hilbert function (or polynomial) guarantees the flatness of M (or of all the $M[f^{-1}]_0$) as long as R_0 is a reduced ring.

Embedded First-Order Deformations

Exercise 6.12: Let k be a field and let $\varphi : S \twoheadrightarrow R = S/I$ be k-algebras. Let A be a k-algebra with a distinguished map $p : A \to k$ such inducing the identity on k. Set $\tilde{S} = A \otimes_k S$. We define an **embedded deformation** of R with **base** A to be an ideal $\tilde{I} \subset \tilde{S}$ such that $\tilde{R} := \tilde{S}/\tilde{I}$ is flat over A and such that $p \otimes 1 : \tilde{S} \to S$ carries $\tilde{I}$ onto I. (Abusing the terminology, we sometimes say that $\tilde{R}$ is an embedded deformation.) For any base algebra A there is at least one embedded deformation, called the **trivial embedded deformation**, obtained by setting $\tilde{I} = A \otimes_k I$.

If R and S are affine domains over k, then the data $S \twoheadrightarrow R$ corresponds to an embedding variety $X \subset Y$. If A is also an affine domain over k, corresponding to a variety Z, and the map $p : A \to k$ corresponds to a point $z \in Z$, then an embedded deformation with base A is a flat family $\tilde{X}$ over Z, embedded in $Y \times Z$,

$$\begin{array}{ccc} \tilde{X} & \subset & Y \times Z \\ \pi \downarrow & \swarrow & \\ Z & & \end{array}$$

such that $\pi^{-1}(z) = X$. The family is **trivial** if $\tilde{X} = X \times Z$ as a subvariety of $Y \times Z$. The notion is useful, for example, when X is singular—one might hope to find a family in which other fibers $\pi^{-1}(z')$ for $z' \in Z$ are smooth.

It is in general quite hard to find embedded deformations, and most of the results about them relate to particular types of rings R or spaces X (determinantal rings, rings with rational singularities, and so forth). Nevertheless, there is a simple way of finding all nontrivial embedded deformations in which the base ring is $k[x]/(x^2)$. We shall describe it here and in Exercise 16.8. We write $k[\varepsilon]$ for $k[x]/(x^2)$, with $\varepsilon^2 = 0$.

Any embedded deformation $A \otimes S \twoheadrightarrow \tilde{R}$ gives rise to embedded deformations over $k[\varepsilon]$, because if $\mathfrak{m}$ is the kernel of the distinguished map $A \to k$, and we choose any map $\mathfrak{m}/\mathfrak{m}^2 \to k$ of k-vector spaces, with kernel $\mathfrak{m}'$, say, then $A/(\mathfrak{m}' + \mathfrak{m}^2) \cong k[\varepsilon]$, and $\tilde{R} \otimes_A A/(\mathfrak{m}' + \mathfrak{m}^2)$ is a deformation over $k[\varepsilon]$. The set of homomorphisms $\mathfrak{m}/\mathfrak{m}^2 \to k$ is by definition the **Zariski tangent space** to A at the maximal ideal $\mathfrak{m}$. (If A corresponds to an affine variety, this is the tangent space at the point corresponding to $\mathfrak{m}$ in a more geometric sense.) We shall see that the set of deformations of R over $k[\varepsilon]$ is naturally a k-vector space. Thus, in a certain natural sense this set is the **Zariski tangent space** to the space of all deformations—even if the latter doesn't exist! See Eisenbud and Harris [1992], Chapter 4, for a more complete view of this idea, which is due to Grothendieck. In some cases there is actually a space of all deformations (the "versal deformation space"); see Schlessinger [1968], Artin [1976], and Sernesi [1986] for a treatment of some of these.

Embedded deformations of $S \twoheadrightarrow R$ over $k[\varepsilon]$ are called **first-order infinitesimal embedded deformations**, in keeping with the idea from the theory of schemes that $k[\varepsilon]$ is the affine ring of the first-order infinitesimal neighborhood of a point in a line.

a. Suppose M is any $k[\varepsilon]$-module. Show that M is flat over $k[\varepsilon]$ iff $(0 :_M \varepsilon)$, the annihilator of ε in M, is equal to εM. (Note that the free resolution of k as a $k[\varepsilon]$-module is

$$\cdots \to k[\varepsilon] \xrightarrow{\varepsilon} k[\varepsilon] \xrightarrow{\varepsilon} k[\varepsilon] \xrightarrow{\varepsilon} \cdots \xrightarrow{\varepsilon} k[\varepsilon] \to k \to 0.$$

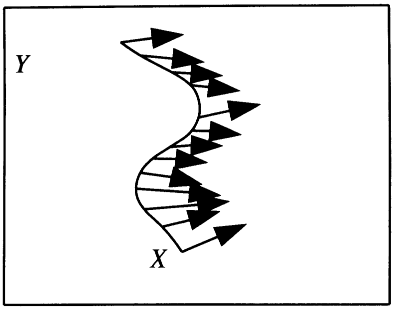


FIGURE 6.9. A normal vector field.

Thus, $(0 :_M \varepsilon)/\varepsilon M = \mathrm{Tor}_1^{k[\varepsilon]}(k, M)$, so the result in question is just the local criterion of flatness, with no hypothesis on M.) Show that this is equivalent to the condition that multiplication by ε from M to εM induces an isomorphism $M/\varepsilon M \to \varepsilon M$.

b. Note that I/I^2, called the **conormal module** of R in S for reasons explained in Chapter 16, is an R-module. Given any homomorphism $\varphi : I/I^2 \to R = S/I$, define an ideal $\tilde{I} \subset S[\varepsilon]$ by $\tilde{I} = \varepsilon I + (1 + \varepsilon\varphi)I$ (here we regard $\varepsilon\varphi$ as a map from I to $S/\varepsilon I$ sending $g \in I$ to $\varepsilon\varphi(g + I^2) \in \varepsilon S/\varepsilon I \subset S/\varepsilon I$. Show that if a set of elements $g_i \in I$ generates I and if g_i' is a representative in S for $\varphi(g_i)$ in S/I, then $\tilde{I}$ is generated by the elements $g_i + \varepsilon g_i'$. Show that $S/\tilde{I}$ is flat over $k[\varepsilon]$, so that we have defined a first-order infinitesimal embedded deformation from an element of the module $N := \mathrm{Hom}(I/I^2, R)$, called the **normal module** of R in S.

c. Given a first-order infinitesimal embedded deformation $\tilde{R} = S[\varepsilon]/\tilde{I}$, we may regard $\tilde{I}$ as a subspace of $S[\varepsilon] = S \oplus S\varepsilon$. As such, show that $\tilde{I}$ projects onto $I \subset S$. It follows that $\tilde{I}$ contains $I\varepsilon \subset S\varepsilon$. Use the flatness of $\tilde{R}$ to show that the image $\tilde{I}/I\varepsilon$ of $\tilde{I}$ in $S \oplus S\varepsilon/I\varepsilon$ is the graph of a homomorphism from I to $S\varepsilon/I\varepsilon \cong S/I = R$; that is, $I/I\varepsilon = \{(g, \psi(g)) | g \in I\}$ for some $\psi \in \mathrm{Hom}(I, R)$. Since I kills R, ψ kills I^2, and thus ψ induces a homomorphism $\varphi \in \mathrm{Hom}(I/I^2, R) = N$.

d. Show that the correspondences defined in parts b and c are inverse to one another. Thus they define a bijection between the set of first-order embedded deformations of $S \twoheadrightarrow R = S/I$ and $N = \mathrm{Hom}_R(I/I^2, R)$. (As we shall see in Chapter 16, if S corresponds to a smooth affine variety Y, and R to a smooth subvariety $X \subset Y$, or a little more generally, then N is the set of **normal vector fields** in Y along X, such as the one in Figure 6.9.) The associated infinitesimal deformation should be thought of as the flow moving each point of X in the direction determined by this vector field.

e. (Compare with Exercise 16.8f) Let $S = k[x]$ and $R = k[x]/(x^n)$ (the "n-fold point on a line"). Show that every first-order infinitesimal embedded deformation may be written in the form

$$S[\varepsilon]/(x^n + a_1\varepsilon x^{n-1} + \cdots + a_n\varepsilon)$$

for a unique $a_1, \ldots, a_n \in k$. (Geometrically, this corresponds to a family of n points on a line approaching 0.)

f. (Compare with Exercise 16.8g) Let $S = k[x, y]$ and $R = k[x]/(xy)$ (the "ordinary double point"). Show that each first-order infinitesimal embedded deformation may be written in the form

$$S[\varepsilon]/(xy + \varepsilon(a + xp(x) + yq(y)))$$

for unique $a \in k$, $p(x) \in k[x]$, $p(y) \in k[y]$. Note that the space of deformations is infinite-dimensional.

g. If R and S are graded, show that N is graded and that the first-order infinitesimal embedded deformations of R as a graded S-algebra correspond to the homogeneous elements of degree 0 in N.

7

Completions and Hensel's Lemma

In this section we shall study the **completion** of a ring R with respect to an ideal $\mathfrak{m}$, written $\hat{R}_{\mathfrak{m}}$, or simply $\hat{R}$ if $\mathfrak{m}$ is clear from the context. The construction is usually applied in the case where R is a local ring and $\mathfrak{m}$ is the maximal ideal. If R is a polynomial ring $R = k[x_1, \ldots, x_n]$ over a field, and $\mathfrak{m} = (x_1, \ldots, x_n)$ is the ideal generated by the variables, then the completion is the ring $k[[x_1, \ldots, x_n]]$ of formal power series over k. More generally, if k is a field and $R = k[x_1, \ldots, x_n]/I$, then the completion of R with respect to $\mathfrak{m} = (x_1, \ldots, x_n)$ is the ring $k[[x_1, \ldots, x_n]]/Ik[[x_1, \ldots, x_n]]$. General completions can similarly be defined in terms of formal power series (Exercise 7.11), but we shall give an intrinsic development.

Completions first appeared in number theory. Hensel worked out and refined the theory of p-adic numbers during a decade or so starting in [1897]. He saw the p-adic numbers as bringing analysis, similar to the local analysis of functions on a Riemann surface, to bear on number theory, and his idea has proved fantastically successful.

7.1 Examples and Definitions

The usefulness of completions can be stated geometrically as follows: A localization R of the affine ring of a variety at the maximal ideal $\mathfrak{m}$ of a point on the variety represents and reflects the properties of Zariski open neighborhoods of the point; the completion $\hat{R}_{\mathfrak{m}}$ represents the properties of the variety in far smaller neighborhoods. For example, over the complex

numbers, the information available from $\hat{R}_{\mathfrak{m}}$ is (roughly speaking) information about arbitrarily small neighborhoods in the "classical topology" induced by the fact that the variety is a closed subspace of some $\mathbf{C}^n$ with its ordinary topology.

A simple example may make this clearer. Consider the two-to-one map π from the parabola to the horizontal line in Figure 7.1.

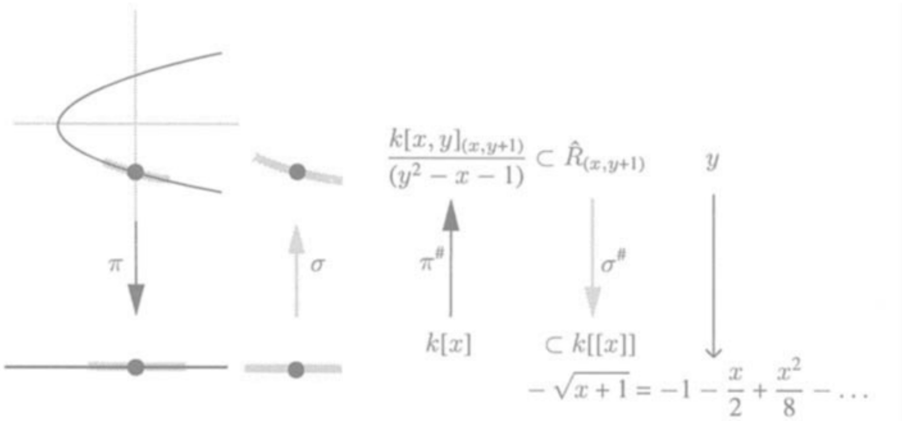


FIGURE 7.1.

Algebraically, the map $\pi^{\#}$ defined by composition with π is the inclusion of the coordinate ring S of the line into the coordinate ring R of the parabola,

$$\pi^{\#} : S = k[x] \to R = k[x, y]/(y^2 - x - 1) \quad x \mapsto x.$$

The derivative of π is nonzero near the point $x = 0, y = -1$ of the parabola. Thus the inverse function theorem tells us (at least in the cases $k = \mathbf{R}$ or $k = \mathbf{C}$) that near the point $x = 0$ on the line there is an analytic function σ from the line to the parabola that is a local inverse to π. But the inverse function theorem fails in algebraic geometry: There is no polynomial mapping σ that is locally the inverse to π, because the element y would have to go to a square root of $x + 1$, and there is no such polynomial. However, $\sqrt{x+1}$ is represented by a power series, so that at the level of power series there is an inverse,

$$\sigma^{\#} : \hat{R}_{(x,y+1)} = k[[x, y]]/(y^2 - x - 1) \to \hat{S}_{(x)} = k[[x]]$$
$$x \mapsto x, y \mapsto -\sqrt{x+1} = -1 - x/2 + x^2/8 - \cdots.$$

If $k = \mathbf{R}$ or $\mathbf{C}$, then this series converges for $|x| < 1$ and represents a function, the inverse of π guaranteed by the inverse function theorem. If k is arbitrary, (of characteristic $\neq 2$) we may still use it as a formal power series. As we shall see, such a thing is generally true for completions and is a variant of the result called Hensel's lemma.

We shall base our treatment of completions on the notion of the **inverse limit**, and we begin by reminding the reader about this useful piece of general algebra. The unproved assertions that follow are quite easy; the reader who has not seen the theory before should prove them as exercises! Appendix 6 contains further information on this construction.

Let R be an abelian group, and let $R = \mathfrak{m}_0 \supset \mathfrak{m}_1 \supset \mathfrak{m}_2 \supset \ldots$ be a sequence of subgroups (a descending filtration). We define the completion $\hat{R}$ of R with respect to the $\mathfrak{m}_i$ to be the inverse limit of the factor groups $R/\mathfrak{m}_i$, which is by definition a subgroup of the direct product:

$$\begin{aligned}\hat{R} &:= \varprojlim R/\mathfrak{m}_i \\ &:= \{g = (g_1, g_2, \ldots) \in \prod\nolimits_i R/\mathfrak{m}_i \mid g_j \equiv g_i (\operatorname{mod} \mathfrak{m}_i) \quad \text{for all } j > i\}.\end{aligned}$$

If R is a ring and all the $\mathfrak{m}_i$ are ideals, then each of the $R/\mathfrak{m}_i$ is a ring, and it follows at once that $\hat{R}$ is also a ring. $\hat{R}$ has a filtration by ideals

$$\hat{\mathfrak{m}}_i := \{g = (g_1, g_2, \ldots) \in \hat{R} \mid g_j = 0 \quad \text{for all } j \leq i\},$$

and it follows at once from the definition that $\hat{R}/\hat{\mathfrak{m}}_i = R/\mathfrak{m}_i$.

The most important case is the one where R is a ring filtered by ideals of the form $\mathfrak{m}_i = \mathfrak{m}^i$ for some ideal $\mathfrak{m}$ of R; this is called the $\mathfrak{m}$-adic filtration of R. The completion of R with respect to $\mathfrak{m}$ is defined to be the completion with respect to the $\mathfrak{m}$-adic filtration. It is denoted by $\hat{R}_{\mathfrak{m}}$. We write $\hat{m}$ for $\hat{m}_1$ in this case. For simplicity, we shall now restrict ourselves to the case of the $\mathfrak{m}$-adic filtration, leaving the easy generalization to the interested reader.

In case $\mathfrak{m}$ is a maximal ideal, we claim that $\hat{R}_{\mathfrak{m}}$ is a local ring with maximal ideal $\hat{m}$. Indeed, $\hat{R}/\hat{\mathfrak{m}}\hat{R}_{\mathfrak{m}} = R/\mathfrak{m}$, a field. Moreover, if $g = (g_1, g_2, \ldots) \in \hat{R}_{\mathfrak{m}} \subset \Pi_i R/\mathfrak{m}^i$ is outside of $\hat{\mathfrak{m}}$, then $g_1 \neq 0$, and it follows that each g_i is outside of $\mathfrak{m}(R/\mathfrak{m}^i)$. Thus each g_i is a unit. From the condition $g_j \equiv g_i (\operatorname{mod} \mathfrak{m}^i)$ for $j > i$ it follows that $g_j^{-1} \equiv g_i^{-1} (\operatorname{mod} \mathfrak{m}^i)$ for $j > i$, so the element $h = (g_1^{-1}, g_2^{-1}, \ldots)$ is in $\hat{R}_{\mathfrak{m}}$ and is the inverse of g. Furthermore, $R/\mathfrak{m}^i$ is equal to its localization $(R/\mathfrak{m}^i)_{\mathfrak{m}} = R_{\mathfrak{m}}/\mathfrak{m}^i_{\mathfrak{m}}$, so we get the same completion by first localizing R and then completing with respect to the localized maximal ideal $\mathfrak{m}_{\mathfrak{m}}$.

Example. If $R = S[x_1, \ldots, x_n]$ is a polynomial ring over the ring S, and $\mathfrak{m} = (x_1, \ldots, x_n)$, then the completion with respect to $\mathfrak{m}$ is the formal power series ring

$$\hat{R}_{\mathfrak{m}} \cong S[[x_1, \ldots, x_n]].$$

Indeed, from the maps $S[[x_1, \ldots, x_n]] \to R/\mathfrak{m}^i$ sending f to $f + \mathfrak{m}^i$, we get a map

$$S[[x_1, \ldots, x_n]] \to \hat{R}_{\mathfrak{m}}$$

sending

$$f \mapsto (f+\mathfrak{m}, f+\mathfrak{m}^2, \ldots) \in \hat{R}_{\mathfrak{m}} \subset \prod R/\mathfrak{m}^i.$$

The inverse map is given by sending $(f_1+\mathfrak{m}, f_2+\mathfrak{m}^2, \ldots) \in \hat{R}_{\mathfrak{m}}$, where the f_i are polynomials and $f_i = f_j +$ (terms of degree $> \min(i,j)$) to the power series $f_1 + (f_2 - f_1) + (f_3 - f_2) + \cdots$. This is a well-defined formal power series because the degree of $f_{i+1} - f_i$ is at least $i+1$, and one checks at once that it is independent of the choice of f_i in $f_i + \mathfrak{m}^i$.

Here is a similar example from number theory, with a subtle difference:

Example. Let $p \in \mathbf{Z}$ be a prime number. The ring $\hat{\mathbf{Z}}_{(p)}$, which is usually written $\mathbf{Z}_p$, is called the **ring of p-adic numbers.**

Elements of this ring, the p-adic numbers themselves, may be written, by a trick like the one above, as power series of the form

$$a_0 + a_1 p + a_2 p^2 + \cdots \quad \text{with } 0 \le a_i < p,$$

but addition is done by "carrying," not "termwise" as in the formal power series ring. For example,

$$(1 + 0p + 0p^2 + \cdots) + ((p-1) + 0p + 0p^2 + \cdots) = 0 + 1p + 0p^2 + \cdots.$$

If you aren't already familiar with the p-adic numbers, you might pause to check, for example, the surprising formula

$$1 + 2 + 4 + 8 + \cdots = -1$$

in the ring of 2-adic integers, by writing out the left-hand side as an element of $\Pi_i \mathbf{Z}/(2^i)$ and adding $1 = (1, 1, 1, \ldots)$.

When the natural map $R \to \hat{R}_{\mathfrak{m}}$ is an isomorphism, we shall say that R is **complete with respect to $\mathfrak{m}$.** When $\mathfrak{m}$ is a maximal ideal, we simply say that R is a **complete local ring.** Note that the ideal $\cap_j \mathfrak{m}^j$ always goes to zero in $\hat{R}_{\mathfrak{m}}$ so that if R is complete with respect to $\mathfrak{m}$, then $\cap_j \mathfrak{m}^j = 0$. This last condition is sometimes expressed by saying that R is **separated with respect to $\mathfrak{m}$** (the terminology is the usual topological one if we give R the "Krull topology," explained later in this chapter).

7.2 The Utility of Completions

There are several reasons why the completion is useful, and we shall describe some of them before giving proofs. First, the completion is closely related to the original ring. For example, it inherits the Noetherian property:

Theorem 7.1. *Let R be a Noetherian ring and let $\mathfrak{m}$ be an ideal of R. Let $\hat{R} = \hat{R}_{\mathfrak{m}}$ be the completion of R with respect to $\mathfrak{m}$.*

a. *$\hat{R}$ is a Noetherian ring.*

b. *$\hat{R}/\mathfrak{m}^j\hat{R} = R/\mathfrak{m}^j$. Thus $\hat{R}$ is complete with respect to $\mathfrak{m}\hat{R}$, and*

$$\mathrm{gr}_{\mathfrak{m}_{\hat{R}}} \hat{R} = \mathrm{gr}_{\mathfrak{m}} R.$$

The following result is one of the main results that help to transmit information between a ring R and its completion.

Theorem 7.2. *Let R be a Noetherian ring and let $\mathfrak{m}$ be an ideal of R. Let $\hat{R} = \hat{R}_{\mathfrak{m}}$ be the completion of R with respect to $\mathfrak{m}$.*

a. *If M is a finitely generated R-module, then the natural map*

$$\hat{R} \otimes_R M \to \varprojlim M/\mathfrak{m}^j M =: \hat{M}$$

is an isomorphism. In particular, if S is a ring that is finite as an R-module, then $\hat{R} \otimes_R S$ is the completion of S with respect to the powers of the ideal $\mathfrak{m}S$.

b. *$\hat{R}$ is flat as an R-module.*

A second reason why the completion is useful is that it is better than the original ring in a crucial respect: Complete rings satisfy Hensel's lemma. The idea is very closely related, as we shall see, to Newton's method for solving equations, and to the implicit function theorem. A special case is suggested in the example at the beginning of this section. Here is the result, which is most often applied in the case where $f'(a)$ is a unit.

Theorem 7.3 (Hensel's Lemma). *Let R be a ring that is complete with respect to the ideal $\mathfrak{m}$, and let $f(x) \in R[x]$ be a polynomial. If a is an approximate root of f in the sense that*

$$f(a) \equiv 0 (\mathrm{mod}\, f'(a)^2\mathfrak{m}),$$

then there is a root b of f near a in the sense that

$$f(b) = 0 \text{ and } b \equiv a (\mathrm{mod}\, f'(a)\mathfrak{m}).$$

If $f'(a)$ is a nonzerodivisor in R, then b is unique.

We shall see that if R is complete with respect to $\mathfrak{m}$ then it is complete with respect to any power of $\mathfrak{m}$, so that Theorem 7.3 handles arbitrary degrees of approximation.

Like Newton's method, Hensel's lemma works for systems of equations in several variables, too. We give a multivariate version of the theorem in Exercise 7.26. Complete rings are not the only ones that satisfy Hensel's lemma. For example, the rings of convergent power series over **R** or **C** also satisfy it. Azumaya [1950] defined a local ring with maximal ideal $\mathfrak{m}$, to be **Henselian** if it satisfies Hensel's lemma. Given a local ring R with maximal ideal $\mathfrak{m}$, there is a smallest ring S containing R and Henselian with respect to $\mathfrak{m}S$; it is called the **Henselization of** $\boldsymbol{R}$ **with respect to** $\boldsymbol{S}$; its existence was proved by Nagata in the 1950s. The Henselization of R is much closer to R than is the completion because it is actually a union of rings finite over R (this is almost never true of the completion). It can thus be used to give the same microscopic view of a variety as the completion, but without passing out of the category of algebraic varieties. See, for example, Milne [1980, Section 1.4] for details and a geometric view.

We shall deduce Theorem 7.3 from Theorem 7.16, which allows us to construct maps from a power series ring to a complete ring, and from Corollary 7.17, which gives a criterion for such a map to be an isomorphism analogous to the inverse function theorem in analytic geometry.

Many statements of Hensel's lemma involve factoring equations; the version given here, in the case where $f'(a)$ is a unit, is just the case where one of the factors is linear. The general factorization result may be deduced from Theorem 7.3 (or even the version in which f is assumed monic) in a page: See, for example, Nagata [1962]. Instead of deriving it this way, we invite the reader to prove it directly in Exercise 7.20. (It can also be deduced from Exercise 7.26 using resultants.) The lifting of idempotents, proved here in Corollary 7.5, is another equivalent version (we show how to deduce the case of Theorem 7.3 in which f is monic from it in Exercise 7.22).

Here are two fairly typical examples of the use of Hensel's Lemma, one from number theory and one from algebraic geometry.

Example (Square roots in the p-adic integers). Which elements $c \in \hat{\mathbf{Z}}_{(p)}$ are perfect squares? Hensel's lemma (together with quadratic reciprocity) can be used to give a complete and easily computable answer to this question.

First of all we may write c uniquely in the form $c = p^n b$ for some non-negative integer n and some element b not divisible by p. Thus c is a square iff n is even and b is a square.

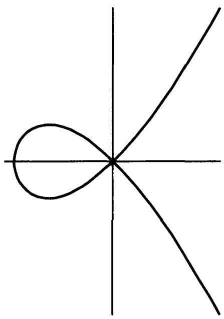
Next we must decide whether b is a square. If $b = a^2$, then reducing mod p we see that the image $\bar{b}$ of b in the field $\hat{\mathbf{Z}}_p/p\hat{\mathbf{Z}}_p = \mathbf{Z}/(p)$ is the square of the image $\bar{a}$ of a. (We could use quadratic reciprocity to check efficiently whether $\bar{b}$ is a square.)

Now consider the polynomial $f(x) = x^2 - b \in \hat{\mathbf{Z}}_p[x]$. Its derivative is $f'(x) = 2x$. If b is a square mod p, say $b \equiv a^2 (\operatorname{mod} p)$, then $\bar{f}(x)$ has a as a root. If $p \neq 2$, then $f'(\bar{a}) = 2(\bar{a}) \neq 0$ in the field $\mathbf{Z}/p\mathbf{Z}$, so we may

apply Hensel's lemma to conclude that b has a p-adic square root. Thus the apparently trivial condition that $\bar{b}$ have a square root is actually a sufficient condition for b to have a square root if $p \neq 2$.

If $p = 2$, then $f'(\bar{a}) = 0$ and the preceding argument fails. However, suppose that $b \equiv 1 (\text{mod}\, 8)$. Then we may take $a = 1$, and we have $f'(a) = 2$, and $f(a) = 1 - b \equiv 0(\text{mod}(2^2 p = 8))$. Thus, Hensel's lemma applies to show that b has a 2-adic square root. The hypothesis $b \equiv 1(\text{mod}\, 8)$ seems restrictive until one notices that if b is a square, then, since b is odd, $b = (1 + 2a)^2 = 1 + 4(a + a^2)$, and 2 divides $a + a^2$, whence $b \equiv 1(\text{mod}\, 8)$. Thus Hensel's lemma gives a complete result in this case too!

Example (Branch of a plane curve). Consider the affine coordinate ring of a nodal plane cubic curve over a field k of characteristic $\neq 2$ (see Figure 7.2).



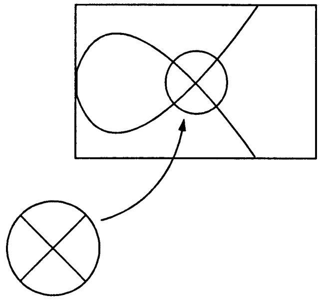
$R = k[x, y]/(y^2 - x^2(1 + x))$

FIGURE 7.2.

As the curve is irreducible, the ring R is a domain, and it follows at once that its localization at the maximal ideal $\mathfrak{m} = (x, y)$, which corresponds to the node, is a domain. This says that every Zariski neighborhood of the node is irreducible—in this case, a Zariski neighborhood consists of the whole curve minus a finite set of points other than the node. (The picture over $\mathbf{R}$, in Figure 7.3, looks like it might possibly become reducible if we leave out a point; but over the complex numbers a neighborhood of the omitted point will be a punctured disk, so the curve remains irreducible.)

However, if we examine a really small neighborhood of the node, either by using convergent power series as functions in the case of the complex numbers or formal power series in general, we see (See Figure 7.3) that in this neighborhood the curve is reducible! This corresponds to the fact that the equation $y^2 - x^2(x + 1)$ can be factored in the power series ring; that is, $x + 1$ has a square root. This follows from Hensel's lemma, exactly as in the earlier example from number theory: The element 1 is a square root of $(x + 1) \,\text{mod}\, x$, and by Hensel's lemma it can be lifted to a square root in the power series ring. Of course, in this case we can write down the square

Spec $k[x,y]_{(x,y)}/(y^2 - x^3 - x^2)$
$\subset$ Spec $k[x,y]_{(x,y)}$
is irreducible;



but its preimage,
corresponding to
$k[[x,y]]/(y^2 - x^3 - x^2)$
is not.

FIGURE 7.3.

root directly, using the Taylor series

$$\sqrt{1+x} = 1 + (1/2)x - (1/8)x^2 + (1/16)x^3 - (5/32)x^4 - \cdots$$

(which is even convergent for $|x| < 1$ over the complex numbers).

Further indications of the importance of Hensel's Lemma can be seen in the following corollaries. The first generalizes the example of the inverse function theorem given in the beginning of this section.

Corollary 7.4. *If $f(t,x)$ is a polynomial in two variables over a field k, and $x = a$ is a simple root of $f(0,x)$, then there is a unique power series $x(t)$ with $x(0) = a$ and $f(t,x(t)) = 0$ identically.*

Proof. Use Theorem 7.3 with $R = k[[t]], \mathfrak{m} = (t)$. □

Since the condition that $f(0,x)$ has a simple root is the condition $\partial f/\partial x$ $(0,a) \neq 0$, this is like the implicit function theorem for polynomials in two variables. More general versions of Hensel's lemma, such as those in the exercises, imitate more general versions of the implicit and inverse function theorems.

7.3 Lifting Idempotents

A striking algebraic consequence of Hensel's lemma is the liftability of idempotents, an idea due to Azumaya [1950]. If A is a (not necessarily commutative) algebra over a commutative ring R, and $e_1, \ldots, e_n \in A$, then we say that the e_i are **idempotent** if $e_i^2 = e_i$. We say that the e_i are **orthogonal idempotents** if in addition, $e_i e_j = e_j e_i = 0$ for $i \neq j$. The elements 0 and 1 are called **trivial idempotents**. Elementary algebra shows that the sum

of any set of orthogonal idempotents is again an idempotent. The set of orthogonal idempotents $\{e_1, \ldots, e_n\}$ is **complete** if $\sum e_i = 1$. If the set is not complete, it can always be completed by adjoining $f := 1 - \sum e_i$; again, elementary algebra shows that f is an idempotent, orthogonal to the e_i.

For example, suppose that M is an R-module with a direct sum decomposition $M = \oplus_{i=1}^{n} M_i$. Let $A = \mathrm{Hom}_R(M, M)$, the endomorphism algebra of M. Let e_i be the projection of M onto its submodule M_i, with kernel $\oplus_{j \neq i} M_j$. Then the e_i form a complete set of orthogonal idempotents of A. Conversely, if $\{e_1, \ldots, e_n\}$ is a complete set of orthogonal idempotents of A, then for any $m \in M$ we have $m = 1m = \sum e_i(m)$, so the $e_i(M)$ together generate M. Furthermore, if $m \in e_i(M) \cap e_j(M)$, then writing $m = e_i(n)$ for some $n \in M$ shows that $e_j(m) = e_j e_i(n) = 0$. But writing $m = e_j(n')$ for some $n' \in M$ shows that $e_j(m) = e_j e_j(n') = e_j(n') = m$, so $m = 0$. Thus $e_i(M) \cap e_j(M) = 0$ for each $i \neq j$, and we see that $M = \oplus e_i(M)$. Thus, complete sets of orthogonal idempotents of A are in one-to-one correspondence with direct sum decompositions of M. In the special case where $M = A$, we write $e_i(A) = e_i A$. If each e_i is in the **center** of A (that is, the set of elements that commute with every element of A)—for example, if A is commutative—then the decomposition $A = \oplus e_i A$ expresses A as a direct product of subalgebras, since $(e_i a)(e_i b) = e_i^2 ab = e_i ab \in e_i A$ for any $a, b \in A$, and $(e_i a)(e_j b) = e_i e_j ab = 0$ for $i \neq j$.

A second important example concerns factorization. Suppose an element $f \in R$ has a factorization $f = gh$ and that g and h are relatively prime in the sense that $(f, g) = R$. Write $1 = ag + bh$ for some $a, b \in R$. Set $A = R/(f)$ and let e_1, e_2 be the images of ag and bh in A. Clearly $e_1 + e_2 = 1$ and $e_1 e_2 = 0$. It follows that $e_1 = e_1 1 = e_1(e_1 + e_2) = e_1^2$, and similarly for e_2, so $\{e_1, e_2\}$ is a complete set of orthogonal idempotents. Thus we have a direct product decomposition $A = e_1 A \times e_2 A$. It is easy to see what these rings are. First note that h annihilates e_1. On the other hand, if an element $p \in A$ annihilates e_1, then $p = (e_1 + e_2)p = e_2 p = bhp$, so p is in the ideal generated by h. Thus $e_1 A = A/(hA) = R/(h)$, and similarly $e_2 A = R/(g)$, so the decomposition in question is $R[x]/(f) = R[x]/(g) \times R[x]/(h)$. (Under some circumstances we can go back from a direct sum decomposition of A to a product decomposition of f; see Exercise 7.22.)

Corollary 7.5. *Let R be a (commutative) Noetherian ring complete with respect to an ideal $\mathfrak{m}$. If A is an R-algebra, possibly not commutative, which is finite as an R-module, then any set of orthogonal idempotents of $A/\mathfrak{m}A$ can be lifted to a set of orthogonal idempotents of A. If A is commutative, then the lifting is unique.*

Proof. Here is the central case: Suppose that $\bar{e} \in R/\mathfrak{m}$ is an idempotent, and let $e \in R$ be any element whose image in $R/\mathfrak{m}$ is $\bar{e}$. Take $f(x)$ to be the polynomial $x^2 - x \in R[x]$ so that the roots of f are precisely the idempotents of R. The derivative $f'(e) = 2e - 1$ is a unit since $(2e - 1)^2 \equiv$

$4e^2 - 4e + 1 \equiv 1 \bmod \mathfrak{m}$, and thus $f'(e)\mathfrak{m} = \mathfrak{m}$. Also, $f(e) \equiv 0 \bmod \mathfrak{m}$. Thus by Hensel's lemma we may find a unique root e_1 of $x^2 - x$ in R that lifts $\bar{e}_1$. (See Exercise 7.23 for a more direct proof.)

To prove the corollary, let $\{\bar{e}_1, \ldots, \bar{e}_n\}$ be a set of orthogonal idempotents of $A/\mathfrak{m}A$. We do induction on n.

First suppose that A is commutative. By Theorem 7.2a, the algebra A is itself complete with respect to $\mathfrak{m}A$, so we may also assume $A = R$. By the central case just treated, there is for each i a unique idempotent $e_i \in R$ lifting $\bar{e}_i$. We must show that these are orthogonal. If $i \neq j$, then $\bar{e}_i\bar{e}_j = 0$, so $e_ie_j \in \mathfrak{m}$. But for any positive integer d we have $e_ie_j = e_i^de_j^d = (e_ie_j)^d \in \mathfrak{m}^d$. Thus, $e_ie_j \in \bigcap_d \mathfrak{m}^d = 0$ as required.

We now drop the hypothesis that A be commutative. If $n = 1$, let e be any element of A that reduces to $\bar{e}_1$. Replacing A by the R-subalgebra generated by e, we reduce to the commutative case.

Next, suppose that $n > 1$ and that the corollary has been proven for sets of at most $n - 1$ orthogonal idempotents. Thus we may find a set of orthogonal idempotents $e_1, \ldots, e_{n-1}$ lifting $\bar{e}_1, \ldots, \bar{e}_{n-1}$. Let e' be any lifting of $\bar{e}_n$. Set $f = 1 - \sum_{i=1}^{n-1} e_i$. Note that $fe_i = e_if = 0$ for all $i \leq n-1$. Further, if $\bar{f}$ denotes the image of f in $A/\mathfrak{m}A$, then $\bar{f}\bar{e}_n = \bar{e}_n\bar{f} = \bar{e}_n$.

Set $e = fe'f$. The element e reduces to $\bar{e}_n$ mod $\mathfrak{m}$ and satisfies $e_ie = ee_i = 0$ for $i \leq n - 1$. We may replace A by the R-subalgebra generated by $e_1, \ldots, e_{n-1}, e$ again reducing to the commutative case. □

The lifting of idempotents can indeed by nonunique in the noncommutative case. See Exercise 7.24.

Because of Corollary 7.5, algebras finite over a complete local ring behave like finite-dimensional algebras over a field. The following result is the extension of Corollary 2.16.

Corollary 7.6. *Let R be a complete local Noetherian ring. If A is a commutative R-algebra that is finite as an R-module, then A has only finitely many maximal ideals $\mathfrak{m}_i$, each localization $A_{\mathfrak{m}_i}$ is a complete local ring finite over R, and $A = \Pi_i A_{\mathfrak{m}_i}$ is the direct product of its localizations.*

Proof. Let $\mathfrak{m}$ be the maximal ideal of R. The hypothesis implies that $A/\mathfrak{m}A$ is a finitely generated module over the field $R/\mathfrak{m}$. By Theorems 2.14 and 2.16, $A/\mathfrak{m}A$ may be written as a product $A/\mathfrak{m}A = \bar{A}_1 \times \ldots \times \bar{A}_n$ of local rings. If $\bar{e}_i$ is the unit element of the subalgebra $\bar{A}_i$, then the $\bar{e}_i$ form a set of orthogonal idempotents of $A/\mathfrak{m}A$. By Corollary 7.5 we may lift them to a set of orthogonal idempotents $\{e_1, \ldots, e_n\}$ of A. Setting $A_i = e_iA$, we see that $A = A_1 \times \ldots \times A_n$. Each A_i is finite over R since it is a direct summand of the R-module A.

If $\mathfrak{n}_i$ is a maximal ideal of A_i, then by Corollary 4.17, $\mathfrak{n}_i \cap R$ is a maximal ideal, so $\mathfrak{n}_i \cap R = \mathfrak{m}$. We see from this that every maximal ideal of A_i contains $\mathfrak{m}A_i$. Since $A_i/\mathfrak{m}A_i$ is a local ring, it follows that A_i is local too, and

$\mathfrak{n}_i$ is its unique maximal ideal. The preimage $\mathfrak{m}_i$ of $\mathfrak{n}_i$ under the projection map $A = \Pi A_i \to A_i$ is a maximal ideal of A. Just as for A_i, the maximal ideals of A must contain $\mathfrak{m}$, so they correspond to the maximal ideals of $A/\mathfrak{m}A$, and are all among the $\mathfrak{m}_i$.

In the localization $A_{\mathfrak{m}_i}$, the idempotent e_i becomes a unit. Since $e_i e_j = 0$ for $j \neq i$, we have $A_{\mathfrak{m}_i} = A_{i\mathfrak{n}_i} = A_i$, and we are done. □

7.4 Cohen Structure Theory and Coefficient Fields

The **Cohen structure theorem** (Cohen [1946]) states, roughly speaking, that any complete local Noetherian ring R is a homomorphic image of a power series ring in finitely many variables over a "nice" ring. If R contains a field, this nice ring may be taken to be a field and the result is another consequence of Hensel's lemma. If R does not contain any field, the nice ring may still be taken to be a complete local principal ideal domain of a special form. Because complete local Noetherian rings are finitely generated in this sense over nice rings, they share certain properties with affine rings, and in this way they are much better behaved than arbitrary Noetherian local rings. We shall prove the structure theorem only for rings that contain fields (these are called **equicharacteristic rings**; see Exercise 7.15 for the reason), but we shall sketch some of the rest.

Theorem 7.7 (Cohen Structure Theorem). *Let R be a complete local Noetherian ring with maximal ideal $\mathfrak{m}$ and residue class field K. If R contains a field, then $R \cong K[[x_1, \ldots, x_n]]/I$ for some n and some ideal I.*

The deepest part of the proof is to show what is obvious from the given isomorphism: that R contains a **coefficient field**, that is, a field that maps isomorphically onto the residue class field $R/\mathfrak{m}$. A more precise statement would be useful: Given a ground field $k \subset R$, it would be nice if there were a coefficient field in R containing k. Such coefficient fields do exist when $R/\mathfrak{m}$ is a (possibly infinite) separable extension of k. We shall not only prove the existence of such coefficient fields, we shall describe them.

For any ring R and maximal ideal $\mathfrak{m} \subset R$ one can ask whether R (or the localization $R_\mathfrak{m}$) contains a coefficient field—that is, a field $K \subset R$ mapping onto the residue class field $R/\mathfrak{m}$. It is not hard to interpret this question geometrically in the case of affine rings over k. In brief, if $X \subset Y$ is an irreducible algebraic subset of an algebraic set Y over a field k, then the local ring of Y along X has a coefficient field containing k if, after perhaps removing a closed set from X and Y, X is a neighborhood retract in Y. In this noncomplete case the question is subtle. For example if k is algebraically closed, $Y = \mathbf{A}_k^n$, and X is a curve, then the coefficient field exists iff X is rational (that is, the field of rational functions on X is isomorphic to $k(t)$). See Exercise 7.18.

To describe the coefficient fields containing a field k in a complete local ring R, we use a notion from general field theory; see Appendix 1 for definitions. If $k \subset K$ are fields, then certain sets of elements of K are called **differential bases for K/k**. The simplest definition uses the K-vector space of differentials $\Omega_{K/k}$, defined in Chapter 16. $\Omega_{K/k}$ is generated by elements $d\alpha$ with $\alpha \in K$. A **differential basis** for K/k is a set of elements $\{\alpha_i\} \subset K$ such that $\{d\alpha_i\} \subset \Omega_{K/k}$ is a vector space basis. We shall use this notion only when K is **separable** over k, a case that includes the case when k is perfect and K/k is an arbitrary extension. In this case differential bases are easy to describe: If k has characteristic 0, then a differential basis is simply a transcendence basis. More generally, if K is separably generated over k (that is, K is a separable algebra extension of a purely transcendental extension of k; this is the case whenever K is separable and finitely generated over k), then a differential basis is a separating transcendence basis. In the general case, in characteristic p, with K separable but perhaps not finitely generated, a differential basis is a **p-basis**: a set of elements $\alpha_i \in K$ such that the monomials in the α_i of degree $< p$ form a vector space basis for K as a $(k * K^p)$-vector space. See Theorem 16.14 for proofs.

If R is a local ring with maximal ideal $\mathfrak{m}$ and $k \subset R$ is a subfield, then since $k - \{0\}$ consists of units, it must be contained in $R - \mathfrak{m}$, and thus k maps isomorphically to a subfield of $R/\mathfrak{m}$. Let $K := R/\mathfrak{m}$. If $B \subset K$ is a subset, then any coefficient field $\tilde{K} \subset R$ contains a unique set $\tilde{B}$ of representatives of the elements of B. If R is complete and B is a differential basis of K/k, then we shall show conversely that there is a unique coefficient field of R containing any set $\tilde{B}$ of representatives for the elements of B.

Theorem 7.8. *Let R be a complete local Noetherian ring with maximal ideal $\mathfrak{m}$ and residue class field K. Suppose that R contains a field k, and that K is separable over k. If B is a differential basis for K over k, then there is a one-to-one correspondence between coefficient fields $\tilde{K} \subset R$ containing k and sets $\tilde{B} \subset R$ of representatives for B, obtained by associating to each $\tilde{K}$ the set $\tilde{B}$ of representatives for B that it contains. If k is perfect of characteristic $p > 0$, then k is contained in every coefficient field of R.*

If R contains any field, then it contains either $\mathbf{Q}$ or $\mathbf{Z}/(p)$ (the quotient field of the image of $\mathbf{Z}$). These fields are all perfect. Since every extension of a perfect field is separable, Theorem 7.8 implies that every complete local ring containing a field contains a coefficient field.

Theorem 7.8 would be false without the hypothesis that K is separable over k; see Exercise 7.17 for an example.

We briefly sketch some of what Cohen proved about an arbitrary complete local ring R; see Cohen [1946], as well as Grothendieck [1961, EGA III 0_{III} 10.3] and [1964, EGA IV, part 1, Section 19], Matsumura [1986], and Bourbaki [1983] for more details.

If the residue class field K of R has characteristic 0, then every integer is invertible in R, so R contains the field $\mathbf{Q}$. The structure of R is then given by Theorem 7.7. If K has characteristic $p > 0$, Cohen showed that there is a complete local domain W whose maximal ideal is generated by p and whose residue class field is K; in case K is perfect, there is a unique such ring $W(K)$ (unique, in fact, up to a unique isomorphism that is the identity on K), which can be given explicitly (it is called the **ring of Witt vectors**, see Serre [1979]). For example, if $K = \mathbf{Z}/(p)$, then $W(K)$ is the ring of p-adic integers $\hat{\mathbf{Z}}_{(p)}$. Returning to our complete local ring R with residue class field K and assuming that the characteristic of K is $p > 0$, Cohen proved that one can write $\hat{R}_\mathfrak{m}$ in the form $W[[x_1, \ldots, x_n]]/I$ for W as above, some n, and some ideal I of the power series ring. The ring $W[[x_1, \ldots, x_n]]$ has many properties in common with a ring of power series of the form $K[[x, x_1, \ldots, x_n]]$—they are both **regular local rings** of dimension $n + 1$. We shall study such rings in Chapter 19.

There is a recent reformulation and extension of a useful part of these results, due to Avramov, Foxby, and Herzog [1994]: If $\varphi : R \to S$ is any homomorphism of local rings sending the maximal ideal $\mathfrak{m}$ of R into the maximal ideal of S, then φ has a factorization $R \to R' \to S$, where R' is complete and local, $R' \to S$ is a surjection, and $R'/\mathfrak{m}R'$ is a regular local ring. Such a factorization is called a **Cohen factorization**. Cohen factorizations are not unique, but any two have a sort of common refinement. In case S is a complete local ring with residue class field of characteristic $p > 0$, for example, the map $\mathbf{Z} \to S$ induces a natural map $\mathbf{Z}_{(p)} \to S$; and we may then take the ring R' to be $W(K)[[x_1, \ldots, x_n]]$, where K is the residue class field of S.

Suppose $X \subset \mathbf{A}^n$ is an affine variety with coordinate ring $A(X)$ and p is a point of X, with $\mathfrak{m}_p$ the maximal ideal of p. Let $R = A(X)_{\mathfrak{m}_p}$ be the localization, the "local ring of X at p", and let $\mathfrak{m}$ be its maximal ideal, the localization of $\mathfrak{m}_p$. The completion $\hat{R}_\mathfrak{m}$ should be thought of as a ring of functions "defined on a very small neighborhood of p." Of course, one consequence of this view is that we would expect $\hat{R}_\mathfrak{m}$ not to have any nilpotent elements (functions with values in a field could hardly be nilpotent!). This is indeed true, though we shall not prove it here.

Theorem 7.9. *If R is a local ring with maximal ideal $\mathfrak{m}$ that is a localization of a ring finitely generated over a field or the ring of integers $\mathbf{Z}$, then the completion $\hat{R}_\mathfrak{m}$ has no nilpotent elements. (See Zariski and Samuel [1958, Vol. II, Chapter 8, Section 13].)*

No such result holds for arbitrary Noetherian rings. In fact a theorem of Larfeldt and Lech [1981] says that if A is any finite-dimensional algebra over a field k (for example $k[x]/(x^2)$), then there is a Noetherian local integral domain R with maximal ideal $\mathfrak{m}$ such that $\hat{R}_\mathfrak{m} \cong A[[x_1, \ldots, x_n]]$ for some n. (For more sophisticated versions, see Heitmann [1993] and the references

there.) This is one of the ways in which the Noetherian property is "too general." There have been attempts to define a more special class of rings that would not only be Noetherian but would also share other good properties of affine rings, such as the one expressed by Theorem 7.9. Nagata's "pseudogeometric" is perhaps the first, and Grothendieck's "excellent" the most recent—perhaps even the definitive—example.

7.5 Basic Properties of Completion

In all of this section R will denote a commutative ring and $\mathfrak{m} \subset R$ will denote an ideal. We will consider the $\mathfrak{m}$-adic filtration $\mathfrak{m}^i$ of R, and the completion $\hat{R} = \hat{R}_{\mathfrak{m}}$. Let $\hat{\mathfrak{m}}_j$ be the kernel of the natural map $\hat{R} \to R/\mathfrak{m}^n$. Thus $\hat{\mathfrak{m}}_n$ consists of all those elements of $\hat{R} \subset \prod_j R/\mathfrak{m}^j$ whose component in $R/\mathfrak{m}^j$ lies in $\mathfrak{m}^n$ for every j (and are thus 0 if $j \leq n$). Note that $\mathfrak{m}^n\hat{R} \subset \hat{\mathfrak{m}}_n \subset (\hat{\mathfrak{m}}_1)^n$; we shall see that in the Noetherian case they are all equal, although in general they may differ (see the example in Bourbaki [1985] Ex. III.2.12.)

Before proving the Theorems above, we explain some useful elementary results. From the definitions we get $\hat{R}/\hat{\mathfrak{m}}_n = R/\mathfrak{m}^n$. It follows that $\hat{R} = \varprojlim \hat{R}/\mathfrak{m}_n$, so $\hat{R}$ is complete with respect to the filtration by the $\hat{\mathfrak{m}}_n$. Further, if we write $\operatorname{gr} \hat{R}$ for the associated graded ring of $\hat{R}$ with respect to this filtration, then the natural map $R \to \hat{R}$ induces an isomorphism $\operatorname{gr}_{\mathfrak{m}} R = \operatorname{gr} \hat{R}$.

It is convenient to define elements of $\hat{R}$ as limits of sequences or series of elements of R. We shall say that a sequence of elements $a_1, a_2, \ldots \in \hat{R}$ **converges** to an element $a \in \hat{R}$ if, for every integer n, there is an integer $i(n)$ so that $a - a_{i(n)} \in \hat{\mathfrak{m}}_n$. It follows that a sequence a_i of elements of $\hat{R}_{\mathfrak{m}}$ converges in $\hat{R}_{\mathfrak{m}}$ iff it is a Cauchy sequence, in the sense that for every integer n there is an integer $i(n)$ such that

$$(*) \qquad a_i - a_j \in \hat{\mathfrak{m}}_n \quad \text{for all } i, j \geq i(n).$$

If the a_i are in R, this condition is clearly the same as the condition

$$a_i - a_j \in \mathfrak{m}^n \quad \text{for all } i, j \geq i(n).$$

A convergent sequence of elements of $\hat{R}$ has a "limit" $a \in \hat{R}$ defined by taking a to be the element of $\Pi_n R/\mathfrak{m}^n$ whose nth coordinate is the same as that of $a_{i(n)}$. We write $a = \lim a_i$. Because the $\hat{\mathfrak{m}}_n$ are ideals, both addition and multiplication are continuous in the sense that if $a = \lim a_i$ and $b = \lim b_i$ then $a_i + b_i$ and $a_i b_i$ are convergent sequences which converge to $a + b$ and ab, respectively. For example, to prove the second equation, if $i(n)$ is chosen so that

$$a_i - a_j \in \hat{\mathfrak{m}}_n \quad \text{for all } i, j \geq i(n)$$

and

$$b_i - b_j \in \hat{\mathfrak{m}}_n \quad \text{for all } i, j \geq i(n),$$

then

$$a_i b_i - a_j b_j = a_i(b_i - b_j) + b_j(a_i - a_j) \in \hat{\mathfrak{m}}_n \quad \text{for all } i, j \geq i(n).$$

Note that condition $(*)$ becomes the usual definition of a Cauchy sequence if, for each $a \in \hat{R}_{\mathfrak{m}}$, we take the sets $a + \hat{\mathfrak{m}}_n$ to be a base for the open neighborhoods of a; the resulting topology is called the **Krull topology**, or the m-adic topology, on R. In fact, the whole theory of completions can be developed on this Cauchy sequence foundation, as the reader will see from Exercises 7.8–7.10.

The simplest way to write down sequences of elements satisfying condition $(*)$ is as the partial sums of a series of elements of R whose ith term is in $\mathfrak{m}^i$.

$$a_i = \sum_{j=1}^{i} b_j, \;\; b_i \in \mathfrak{m}^i.$$

In this case we define the infinite sum $\sum_{j=1}^{\infty} b_j$ to be the limit of the a_i. Note that this is exactly what we do when we write down formal power series—in that case the ideal $\mathfrak{m}$ is generated by the variables of a polynomial ring.

We can use these ideas, for example, to make sense of the usual Taylor formula for $1/(1+x)$ in the context of complete rings.

Proposition 7.10. *If R is complete with respect to an ideal $\mathfrak{m}$, then the elements of the multiplicatively closed set $U := \{1 + a | a \in \mathfrak{m}\}$ are units in R.*

Proof. If $a \in \mathfrak{m}$, then $b = 1 - a + a^2 - \ldots$ is a power series that converges in $\hat{R}$; the product $(1+a)b$ is the limit of the series

$$(1+a) - (1+a)a + (1+a)a^2 - \cdots.$$

The ith partial sum of this series is $1 + a^i$, so the series converges to 1. □

Corollary 7.11. *If R is a local ring with maximal ideal P, then the power series ring*

$$R[[x_1, \ldots, x_n]]$$

in indeterminates x_i is a local ring with maximal ideal

$$P + (x_1, \ldots, x_n).$$

Proof. An element f outside $P + (x_1, \ldots, x_n)$ has constant term u outside of P—thus a unit of R. The element $u^{-1}f$ is of the form $1 + g(x)$, with $g(x) \in (x_1, \ldots, x_n)$. Thus, $u^{-1}f$ is a unit, so f is too. □

Given this result, it is interesting to ask which localization we get by inverting the elements of $U := \{1 + a | a \in \mathfrak{m}\}$. See Exercise 7.3.

Next we present a more serious application of the idea of convergence.

Proposition 7.12. *Suppose that R is a ring that is complete with respect to a filtration by ideals $\mathfrak{m}_i$. Let gr R be the associated graded ring of R with respect to this filtration, and for $a \in R$ let* $\mathrm{in}(a)$ *denote the initial form with respect to this filtration. Suppose that $I \subset R$ is an ideal, and $a_1, \ldots, a_s \in I$. If* $\mathrm{in}(a_1) \ldots, \mathrm{in}(a_s)$ *generate* $\mathrm{in}(I)$ *as an ideal in gr R, then $a_1, \ldots, a_s$ generate I.*

Proof. Let $I' = (a_1, \ldots, a_s)$. We may as well assume that no $a_i = 0$, and then we may choose a number d so large that none of the a_i is contained in $\mathfrak{m}_d$. Given any $f \in I$, with $\mathrm{in}(f)$ of degree e, say, we may write

$$\mathrm{in}(f) = \sum_j G_j \,\mathrm{in}(a_j),$$

with $G_j \in \mathrm{gr}_\mathfrak{m} R$ homogeneous of degree equal to $\deg(\mathrm{in}(f)) - \deg(\mathrm{in}(a_j))$.

It follows that if we choose $g_j \in R$ with $\mathrm{in}(g_j) = G_j$ then $f - \sum_j g_j a_j$ lies in $\mathfrak{m}_{e+1}$. Repeating this procedure, we eventually get an element $f' \in I'$ such that $f - f' \in \mathfrak{m}_{d+1}$. Thus we may assume from the outset that $f \in \mathfrak{m}_{d+1}$.

Under these circumstances, the G_j defined above are of degree greater than or equal to $e - d > 0$, and thus we may take $g_j \in \mathfrak{m}_{e-d}$. Now repeating this procedure, we define elements $g_j^{(i)} \in \mathfrak{m}_{e-d+i}$ such that

$$\begin{aligned} f - \sum_j g_j^{(0)} a_j - \sum_j g_j^{(1)} a_j - \sum_j g_j^{(2)} a_j - \cdots - \sum_j g_j^{(n)} a_j \\ = f - \sum_j \sum_{i=0}^{n} g_j^{(i)} a_j \in \mathfrak{m}_{e+n+1}. \end{aligned}$$

The series $\sum_{i=1}^{\infty} g_j^{(i)}$ converges in R; we write h_j for its limit. Because limits preserve finite sums and products, we get $f = \sum_j h_j a_j \in I'$ as required. □

Proof of Theorem 7.1a. With notation as in Proposition 7.12, $\mathrm{gr}\, \hat{R}_{\hat{\mathfrak{m}}} = \mathrm{gr}_\mathfrak{m} R$. Since R is Noetherian, $R/\mathfrak{m}$ is Noetherian. The ring $\mathrm{gr}_\mathfrak{m} R$ is generated over $R/\mathfrak{m}$ by a basis for the finite-dimensional vector space $\mathfrak{m}/\mathfrak{m}^2$, so $\mathrm{gr}_\mathfrak{m} R$ is Noetherian by the Hilbert basis theorem (Theorem 1.2). Thus for any ideal $I \subset \hat{R}_\mathfrak{m}$ the ideal $\mathrm{in}(I)$ is generated by the initial forms of finitely many elements $a_1, \ldots, a_s \in I$. It follows from Proposition 7.12 that the a_i generate I, so I is finitely generated. □

Another consequence suffices to prove Theorem 7.1b.

Corollary 7.13. *If R is Noetherian, then $\hat{\mathfrak{m}}_n = \mathfrak{m}^n \hat{R}_{\mathfrak{m}}$. In particular, $\hat{R}/\mathfrak{m}^j\hat{R} = R/\mathfrak{m}^j$, $\hat{R}$ is complete with respect to $\mathfrak{m}\hat{R}$, and*

$$\mathrm{gr}_{\mathfrak{m}_{\hat{R}}} \hat{R}_{\mathfrak{m}} = \mathrm{gr}_{\mathfrak{m}} R.$$

Proof. The inclusion $R \subset \hat{R}_{\mathfrak{m}}$ induces an isomorphism

$$\mathrm{gr}\, \hat{R}_{\mathfrak{m}} = \mathrm{gr}_{\mathfrak{m}} R,$$

where $\mathrm{gr}\ \hat{R}_{\mathfrak{m}}$ is the graded ring with respect to the filtration by the ideals $\mathfrak{m}_i$.

For the first statement, it suffices by Proposition 7.12 to show that the two ideals have the same initial ideals in $\mathrm{gr}\ \hat{R}_{\mathfrak{m}}$, and this is obvious because both initial ideals consist of all elements of degree $\geq n$. The other statements follow at once. □

We now turn to the question of flatness. We first need a criterion for two filtrations (of an abelian group, say) to give the same completion. It is enough for the two filtrations to be comparable in a sense made precise in the following lemma.

Let $R = \mathfrak{n}_0 \supset \mathfrak{n}_1 \supset \ldots$ be another filtration of R and write $\hat{R}_{\mathfrak{m}}$ and $\hat{R}_{\mathfrak{n}}$ for the respective completions.

Lemma 7.14. *If for each $\mathfrak{n}_j$ there is an $\mathfrak{m}_i$ such that $\mathfrak{m}_i \subset \mathfrak{n}_j$, and for each $\mathfrak{m}_j$ there is an $\mathfrak{n}_i$ such that $\mathfrak{n}_i \subset \mathfrak{m}_j$, then there is a natural isomorphism $\hat{R}_{\mathfrak{m}} \cong \hat{R}_{\mathfrak{n}}$.*

Proof. First, suppose that the $\mathfrak{n}_j$ are simply a subset of the $\mathfrak{m}_i$; the condition of the lemma says in this case that infinitely many $\mathfrak{m}_i$ are among the $\mathfrak{n}_j$. In this case the projection onto a subproduct

$$\prod_i R/\mathfrak{m}_i \to \prod_j R/\mathfrak{n}_j$$

clearly induces a natural isomorphism $\hat{R}_{\mathfrak{m}} \cong \hat{R}_{\mathfrak{n}}$, so we are done.

In the general case we may choose injective functions $\alpha, \beta, \gamma : \mathbf{N} \to \mathbf{N}$ such that

$$\mathfrak{m}_j \supset \mathfrak{n}_{\alpha(j)} \supset \mathfrak{m}_{\beta(j)} \supset \mathfrak{n}_{\gamma(j)},$$

and these induce maps $R/\mathfrak{n}_{\gamma(j)} \to R/\mathfrak{m}_{\beta(j)} \to R/\mathfrak{n}_{\alpha(j)} \to R/\mathfrak{m}_j$ for all j, and thus natural maps as in Figure 7.4.

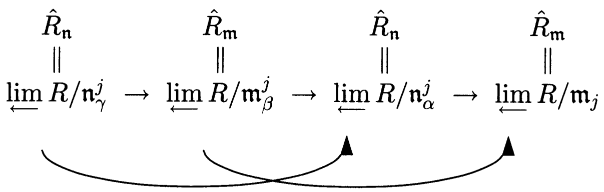


FIGURE 7.4.

Since the maps from the first to the third terms and from the second to the fourth terms are the isomorphisms treated in the special case above, we are done. □

The next step is to show that in suitable circumstances completions preserve exact sequences. This is very close to Theorem 7.2 and is a key result in making completions usable. The nontrivial part of the proof is a telling application of the Artin-Rees lemma. This is the most subtle step in the theory; in general taking limits is not right exact. See Exercise A6.11.

Lemma 7.15. *Let R be a Noetherian ring and let $\mathfrak{m}$ be an ideal of R. If*

$$0 \to A \to B \to C \to 0$$

is a short exact sequence of finitely generated R-modules, then

$$0 \to \varprojlim A/\mathfrak{m}^j A \to \varprojlim B/\mathfrak{m}^j B \to \varprojlim C/\mathfrak{m}^j C \to 0$$

is exact. Thus, completion with respect to the $\mathfrak{m}$-adic filtration preserves exact sequences of finitely generated modules.

Proof. The second statement follows from the first because any exact sequence

$$\cdots \to A_{n+1} \xrightarrow{\varphi_{n+1}} A_n \xrightarrow{\varphi_n} A_{n-1} \to \cdots$$

can be written as a "composite" of short exact sequences

$$0 \to \operatorname{im} \varphi_{n+1} \to A_n \to \operatorname{im} \varphi_n \to 0.$$

To prove the first statement, we begin by showing that $\varprojlim B/\mathfrak{m}^j B \to \varprojlim C/\mathfrak{m}^j C$ is an epimorphism. This follows by an easy diagram chase: If $(c_j + \mathfrak{m}^j C) \in \varprojlim C/\mathfrak{m}^j C$, then we must show that there is an element $(b_j + \mathfrak{m}^j B) \in \varprojlim B/\mathfrak{m}^j B$ mapping to $(c_j + \mathfrak{m}^j C)$. That is, we must show that there is a sequence of elements $b_j \in B$ such that

a. $b_j \mapsto c_j (\operatorname{mod} \mathfrak{m}^j C)$, and

b. $b_j \equiv b_i (\operatorname{mod} \mathfrak{m}^i B)$ if $i < j$; it is of course enough to check this for $i = j - 1$.

We choose these inductively: Having chosen $b_1, \ldots, b_j$ satisfying conditions a and b, we choose an arbitrary element b'_{j+1} mapping to $c_{j+1} \bmod \mathfrak{m}^{j+1}C$. Both b'_{j+1} and b_j map to the same element $c_j \bmod \mathfrak{m}^j C$. But the sequence

$$A/\mathfrak{m}^j A \to B/\mathfrak{m}^j B \to C/\mathfrak{m}^j C \to 0$$

is exact since it is $R/\mathfrak{m}^j \otimes \{A \to B \to C \to 0\}$; so there is an element $a_{j+1} \in A$ such that $a_{j+1} \mapsto b_j - b'_{j+1} \bmod \mathfrak{m}^j$. The element $b_{j+1} := b'_{j+1} + a_{j+1}$ satisfies both conditions.

It remains to show that

$$(*) \qquad 0 \to \varprojlim A/\mathfrak{m}^j A \to \varprojlim B/\mathfrak{m}^j B \to \varprojlim C/\mathfrak{m}^j C$$

is exact. To do this we wish to replace $\varprojlim A/\mathfrak{m}^j A$ by $\varprojlim A/(A \cap \mathfrak{m}^j B)$, because $*$ is then replaced by the limit of the exact sequences

$$0 \to A/(A \cap \mathfrak{m}^j B) \to B/\mathfrak{m}^j B \to C/\mathfrak{m}^j C,$$

and it is easy to show that such a limit is exact (that is, the inverse limit of left-exact sequences is left-exact).

To make the replacement, we must show that the filtration of A by the submodules $\mathfrak{m}^j A$ gives the same completion as the filtration by the submodules $A \cap \mathfrak{m}^j B$. Indeed, it is clear that $A \cap \mathfrak{m}^j B \supset \mathfrak{m}^j A$, and by the Artin-Rees lemma (Lemma 5.1) there is a number k such that $A \cap \mathfrak{m}^j B = \mathfrak{m}^{j-k}(A \cap \mathfrak{m}^k B) \subset \mathfrak{m}^{j-k} A$. By the criterion of Lemma 7.14, the two filtrations give the same completion, so the desired replacement is legitimate.

Now we must show that

$$0 \to \varprojlim A/A \cap \mathfrak{m}^j B \to \varprojlim B/\mathfrak{m}^j B \to \varprojlim C/\mathfrak{m}^j C$$

is left-exact. This follows directly from the definition of the inverse limit. If

$$(b_1, b_2, \ldots) \in \varprojlim B/\mathfrak{m}^j B \text{ goes to } (0, 0, \ldots) \in \varprojlim C/\mathfrak{m}^j C,$$

then each b_j goes to 0 in $C/\mathfrak{m}^j C$. Thus $b_j \in A/A \cap \mathfrak{m}^j B$, and

$$(b_1, b_2, \ldots) \in \varprojlim A/A \cap \mathfrak{m}^j B$$

as required. □

Proof of Theorem 7.2a. The second statement follows at once from the first.

To prove the first, we begin with the case $M = R$, where the result is simply the definition of $\hat{R} = \hat{R}_{\mathfrak{m}}$. It follows at once from the definition that $\varprojlim$ commutes with finite-direct sums. Thus the result is true for finitely generated free modules. Now let M be any finitely generated module, and let

$$F \to G \to M \to 0$$

be a free presentation of M. From the fact that $\varprojlim$ preserves the exactness of sequences of finitely generated modules, it follows that the top row in

the diagram

$$\begin{array}{ccccccc}
\hat{F} & \rightarrow & \hat{G} & \rightarrow & \hat{M} & \rightarrow & 0 \\
\uparrow & & \uparrow & & \uparrow & & \\
\hat{R} \otimes_R F & \rightarrow & \hat{R} \otimes_R G & \rightarrow & \hat{R} \otimes_R M & \rightarrow & 0
\end{array}$$

is right-exact. Of course, the bottom row is right-exact by the right-exactness of tensor products, and the two vertical maps on the left are isomorphisms by what we have already proved. A diagram chase now shows that the right-hand vertical map is also an isomorphism, as required. □

7.2b. By Proposition 6.1 it is enough to show that the multiplication map

$$I \otimes_R \hat{R} \rightarrow I\hat{R} \subset \hat{R}$$

is a monomorphism for finitely generated ideals. By part a, this is the same as showing that the map

$$\hat{I} \rightarrow \hat{R}$$

is a monomorphism. This follows from Lemma 7.15.

7.6 Maps from Power Series Rings

Our next results concern homomorphisms of complete rings. One of the things that makes a polynomial algebra $R[x_1, \ldots, x_n]$ nice is that a map from it to another R-algebra S may be specified by simply telling where to send each of the x_i. The power series ring has a similar property, but only with respect to complete rings S.

Theorem 7.16. *Let R be any ring and let S be an R-algebra that is complete with respect to some ideal $\mathfrak{n}$. Given $f_1, \ldots, f_n \in \mathfrak{n}$:*

a. *There is a unique R-algebra homomorphism*

$$\varphi : R[[x_1, \ldots, x_n]] \rightarrow S$$

sending x_i to f_i for each i. The map φ takes a power series $g(x_1, \ldots, x_n)$ to $g(f_1, \ldots, f_n) \in S$.

b. *If the induced map $R \rightarrow S/\mathfrak{n}$ is an epimorphism and $f_1, \ldots, f_n$ generate $\mathfrak{n}$, then φ is an epimorphism.*

c. *If the induced map of associated graded rings*

$$\operatorname{gr} \varphi : R[x_1, \ldots, x_n] \cong \operatorname{gr}_{(x_1,\ldots,x_n)} R[[x_1, \ldots, x_n]] \rightarrow \operatorname{gr}_{\mathfrak{n}} S$$

is a monomorphism, then φ is a monomorphism.

Note that part b would follow at once from Nakayama's lemma if we knew in advance that S was a finitely generated module over $R[[x_1, \ldots, x_n]]$. In fact, the hypothesis of Nakayama's lemma can be weakened, in the case of complete rings, to include this case. See Exercise 7.2.

Proof.

a. The unique R-algebra map $R[x_1, \ldots, x_n] \to S/\mathfrak{n}^t$ sending x_i to the class of f_i factors through

$$R[[x_1, \ldots, x_n]]/(x_1, \ldots, x_n)^t = R[x_1, \ldots, x_n]/(x_1, \ldots, x_n)^t \to S/\mathfrak{n}^t$$

and thus induces unique maps $R[[x_1, \ldots, x_n]] \to S/\mathfrak{n}^t$ sending x_i to the class of f_i. Since S is the inverse limit of the $S/\mathfrak{n}^t$, there is a unique map $\varphi : R[[x_1, \ldots, x_n]] \to S$ sending x_i to f_i, as required. The image of $g+(x_1, \ldots, x_n)^t$ in $S/\mathfrak{n}^t$ is $g(f_1, \ldots, f_n)+\mathfrak{n}^t$ for every t, so the image of g in S is $g(f_1, \ldots, f_n)$, which makes sense precisely because S is complete with respect to $\mathfrak{n}$.

b. It follows from our hypothesis that the map

$$(x_1, \ldots, x_n)/(x_1, \ldots, x_n)^2 \to \mathfrak{n}/\mathfrak{n}^2$$

is a surjection, so the induced map

$$\operatorname{gr} \varphi : \operatorname{gr}_{(x_1, \ldots, x_n)} R \to \operatorname{gr}_{\mathfrak{n}} S$$

is also a surjection. Now, given $0 \neq g \in S$, let i be the largest number such that $g \in \mathfrak{n}^i$—such an i exists because S is complete, so $\cap \mathfrak{n}^j = 0$. Since $\operatorname{gr} \varphi$ is a surjection we may find an $g_1 \in (x_1, \ldots, x_n)^i$ whose initial form is carried to the initial form of g. It follows that $g - \varphi(f_1) \in \mathfrak{n}^{i+1}$.

Iterating this process, we obtain a sequence of elements $g_j \in (x_1, \ldots, x_n)^{i+j}$ such that $g = \sum_{j=1}^{\infty} \varphi(g_j)$. Because φ preserves infinite sums, this yields $g = \varphi(\sum_{j=1}^{\infty} g_j)$, and we are done.

c. If $0 \neq g \in R[[x_1, \ldots, x_n]]$, then $\operatorname{in}(g)$ is a nonzero form, say of degree d, and from our hypothesis we get

$$\operatorname{gr} \varphi(\operatorname{in}(g)) \neq 0$$

in the degree d part of $\operatorname{gr}_{\mathfrak{n}} S$. But $g \equiv \operatorname{in}(g) \operatorname{mod} (x_1, \ldots, x_n)^{d+1}$, so $\varphi(g) \equiv \operatorname{gr} \varphi(\operatorname{in}(g)) \operatorname{mod} \mathfrak{n}^{d+1}$, whence $\varphi(g) \neq 0$ as well. □

To exploit the Theorem we introduce some notation. If $f \in R[[x]]$ is a power series in one variable, we write $f'(x)$ for the result of differentiating f term by term with respect to x. Thus for example $f(x) = f(0) + f'(0)x+$ (higher order terms).

Corollary 7.17. *Let $f \in xR[[x]]$ be a power series. If φ is the endomorphism*

$$\varphi : R[[x]] \to R[[x]]; \quad x \mapsto f,$$

which is the identity on R and sends x to f, then φ is an isomorphism iff $f'(0)$ is a unit in R.

Proof. Suppose that φ is an isomorphism. The elements of $R[[x]]$ not in (x) are those with nonzero constant term, and φ preserves this subset. Since φ is an isomorphism, it follows that $\varphi((x)) = (x)$. In particular, the image $\varphi(x) = f$ of the generator X of (x) is a generator of (x). We deduce that $f + (x^2)$ generates $(x)/(x^2)$. Since $f \equiv f'(0)x(\text{mod}(x^2))$, we see that $f'(0)$ is a unit of R.

Conversely, suppose that $f'(0) = u$ is a unit of R. We have $\text{gr}_{(x)} R[[x]] = R[x]$, and

$$\text{gr}\, \varphi : R[x] \to R[x]; \quad x \mapsto ux,$$

is an isomorphism because u is a unit. By Theorem 7.16 φ is injective. We may write $f = ux + hx^2 = (u + hx)x$ for some $h \in R[[x]]$. Since $u + hx$ is a unit in $R[[x]]$, we see that f generates (x). Again by Theorem 7.16, φ is surjective, and thus an isomorphism. □

We can use this to prove Hensel's lemma.

Proof of Theorem 7.3. To simplify notation, set $f'(a) = e$. We may choose $h(x)$ so that

$$\begin{aligned} f(a + ex) &= f(a) + f'(a)ex + h(x)(ex)^2 \\ &= f(a) + e^2(x + x^2h(x)). \end{aligned}$$

By Theorem 7.16 there is a ring homomorphism $\varphi : R[[x]] \to R[[x]]$ that is the identity on R and takes x to $x + x^2h(x)$. By Corollary 7.17, φ is an isomorphism. Applying φ^{-1} to the above equation, we obtain

$$f(a + e\varphi^{-1}(x)) = f(a) + e^2x.$$

By hypothesis, we may write $f(a) = e^2c$ with $c \in \mathfrak{m}$. By Theorem 7.16 there is an algebra homomorphism ψ that is the identity on R and carries x to $-c$. Applying it, we get

$$f(a + e\psi\varphi^{-1}(x)) = 0,$$

so $b = a + e\psi\varphi^{-1}(x)$ is the desired element.

Suppose now that e is a nonzerodivisor. To prove the uniqueness of b, suppose that both b and b_1 are roots of f differing from a by elements of $e\mathfrak{m}$, say $b = a + er$ and $b_1 = a + er_1$, with $r, r_1 \in \mathfrak{m}$. By Theorem 7.16 there are ring homomorphisms $\beta, \beta_1 : R[[x]] \to R[[x]]$ that are the identity on R and take x to r and to r_1 respectively. Applying them to the above

formulas produces

$$0 = f(a + er) = f(a) + e^2(r + r^2h(r)),$$
$$0 = f(a + er_1) = f(a) + e^2(r_1 + r_1^2h(r_1)).$$

Subtracting and using the assumption that e is a nonzerodivisor we see that $r+r^2h(r) = r_1+r_1^2h(r_1)$, that is, $\beta\varphi(x) = \beta_1\varphi(x)$. By the uniqueness statement of Theorem 7.16, we get $\beta\varphi = \beta_1\varphi$, and since φ is an isomorphism, $\beta = \beta_1$. Thus $r = r_1$ as required. □

We now turn to the proof of the Cohen structure theorem. First we deal with the existence of coefficient fields. We shall give separate proofs for each of two overlapping cases. We first treat the case where the residue class field has a separating transcendence basis (this includes all cases in characteristic 0 and the finitely generated case in characteristic p). The only tool that is necessary here is Hensel's lemma. Next we treat the case where the residue class field has characteristic p. Here the coefficient field can be described using the pth power map.

Proof of Theorem 7.8. Let R be any local ring containing a field k, and let K be the residue class field of R. If B is a subset of K algebraically independent over k, and $\tilde{B}$ is any set of representatives for B, then every nontrivial polynomial in the elements of $\tilde{B}$ with coefficients in K has nonzero image in K, and is thus invertible in R. It follows that R contains the field $k(\tilde{B})$ of rational functions in the elements of $\tilde{B}$, and this field maps isomorphically to $k(B)$.

Now suppose that K is separable over k and that B is a differential basis for K/k. In characteristic 0 the hypothesis is equivalent to B being a transcendence basis for K/k; in the case of characteristic $p > 0$, Theorem A1.3c shows that the elements of B are algebraically independent over $k * K^{p^\infty}$. In either case, if $\tilde{B} \subset R$ is a set of representatives for B, then the field $k(\tilde{B})$ is contained in R. Under the hypothesis that R is complete and Noetherian, we shall show that there is a unique coefficient field $\tilde{K}$ for R containing $k(\tilde{B})$.

Suppose first that the characteristic of k is 0, or more generally that K is separably algebraic over $k(B)$. By Zorn's lemma, we may choose a subfield K' of K, containing $k(B)$ and maximal among subfields containing $k(B)$ that have unique liftings to subfields of R containing $k(\tilde{B})$. Let $\tilde{K'} \subset R$ be its lifting. We must show that $K' = K$. Let $a \in K$ be an element, and let $f(t)$ be the monic irreducible polynomial with coefficients in K' such that $f(a) = 0$. Using the inverse of the isomorphism $\tilde{K'} \to K'$, we may lift f to a monic polynomial $\tilde{f}$ with coefficients in R. Since K' is separable over K the roots of f are distinct, so the derivative $f'(a) \neq 0$ in K. By Hensel's lemma (Theorem 7.3), there is a unique root $\tilde{a} \in R$ of $\tilde{f}$ whose image in K is a. The field $\tilde{K'}(\tilde{a})$ is thus the unique field lifting $K'(a)$ and containing

$\tilde{K}'$. Putting this together with the uniqueness of $\tilde{K}'$, we see that $\tilde{K}'(\tilde{a})$ is in fact the unique field lifting $K'(a)$ and containing $k(\tilde{B})$. Since K' was maximal, we must have $a \in K'$; that is, $K' = K$ as required.

Now consider the general case where the characteristic of k is $p > 0$. We shall show that

$$\tilde{K} := \bigcap_{q=p^n, n\geq 1} k * R^q[\tilde{B}]$$

is the unique coefficient field of R containing k and $\tilde{B}$. Here R^q denotes the ring of qth powers of elements of R, and $k * R^q[B]$ is the smallest subring of R containing k, R^q, and B. If k' is a perfect field contained in R, then $k' = k'^q \subset R^q$ for every $q = p^n$, so that $k' \subset \tilde{K}$. This proves the last statement of the theorem as well.

We first show that any coefficient field $K' \subset R$ that contains k and $\tilde{B}$ must be contained in $\tilde{K}$. To see this, note that since $K' \cong K$ by an isomorphism carrying $\tilde{B}$ to B, the set $\tilde{B}$ is a p-basis for K' over k. Thus by Theorem A1.4a, we have $K' = k * K'^q[\tilde{B}] \subset k * R^q[\tilde{B}]$ for every $q = p^n$, as required.

Next we define a homomorphism $\varphi : K \to \tilde{K} \subset R$. For $a \in K$ and for each $q = p^n$, let a_q be a representative of a in $k * R^q[\tilde{B}]$; such a representative exists because $k * K^q[B] = K$ by Theorem A1.4a. If a'_q is another such representative, we claim that $a_q - a'_q \in \mathfrak{m}^q$, where $\mathfrak{m}$ is the maximal ideal of R. Once this is established, it follows that the sequence $a_1, a_p, a_{p^2}, \ldots$ converges in R to a limit $\tilde{a} \in \cap_{q=p^n,n\geq 1} k * R^q[\tilde{B}] = \tilde{K}$, independent of the representatives a_q chosen. We set $\varphi(a) = \tilde{a}$. If $r \in \tilde{K}$ and r has image $a \in K$, then we may take $a_q = r$ for all q, so $\varphi(a) = r$. Thus $\tilde{K} = \varphi(K)$. Of course, the image of $\varphi(a)$ in K is just a. The independence of choices shows immediately that $\varphi(a + b) = \varphi(a) + \varphi(b)$ and similarly for multiplication, so φ is a homomorphism and $\tilde{K}$ is the unique coefficient field containing $\tilde{B}$.

It remains to show that with notation as above, $a_q - a'_q \in \mathfrak{m}^q$. By definition a_q and a'_q are polynomials in the elements of $\tilde{B}$ with coefficients in $k * R^q$. Any qth power of an element of $\tilde{B}$ can be absorbed into the coefficients, so we may write

$$a_q = \sum_{w\in W} u_w r_w^q w, \quad a'_q = \sum_{w\in W} u'_w r_w'^q w,$$

with $u_w, u'_w \in k$ and $r_w, r'_w \in R$, where W is the set of monomials in some $b_1, \ldots, b_s \in B$, with degree $< q$ in each b_i. Since $a_q - a'_q \in \mathfrak{m}$, the fact that W is a basis for $k * K^q[b_1, \ldots, b_s]$ as $k * K^q$ vector space shows that $u_w r_w^q - u'_w r_w'^q \in \mathfrak{m}$ for every $w \in W$.

Let $\bar{r}_w$ and $\bar{r}'_w$ be the images of r_w and r'_w in K. Since $\bar{u}_w/\bar{u}'_w = (\bar{r}'_w/\bar{r}_w)^q \in k^q$, and K contains no nontrivial purely inseparable extensions of k, we must have $(\bar{r}'_w/\bar{r}_w) = v \in k$. Thus

$$\begin{aligned}(1/u'_w)(u_w r_w^q - u'_w r_w'^q) &= v^q r_w^q - r_w'^q \\ &= (v r_w - r'_w)^q \in \mathfrak{m}.\end{aligned}$$

Since $\mathfrak{m}$ is prime, $(vr_w - r'_w) \in \mathfrak{m}$, and thus $(vr_w - r'_w)^q \in \mathfrak{m}^q$. Multiplying by u'_w we see that $(u_w r_w^q - u'_w r'^q_w) \in \mathfrak{m}^q$. From this we get $a_q - a'_q \in \mathfrak{m}^q$, as required. □

Proof of Theorem 7.7. Choose a coefficient field $K \subset R$, and let $a_1, \ldots, a_n$ be a set of generators for the maximal ideal of R. Since R is complete, Theorem 7.16a shows that there is a map $\varphi : K[[x_1, \ldots, x_n]] \to R$ sending x_i to a_i. Theorem 7.16b shows that this map is a surjection, so that if $I = \ker \varphi$ then $R \cong K[[x_1, \ldots, x_n]]/I$. □

7.7 Exercises

Exercise 7.1: Let $\mathfrak{m}$ be a maximal ideal of a ring R. Show that the map $R \to \hat{R}_\mathfrak{m}$ factors through the localization map $R \to R_\mathfrak{m}$.

Exercise 7.2: Suppose that M is a module over a ring R that is complete with respect to an ideal $\mathfrak{m}$. We say that M is **separated** if $\cap_k \mathfrak{m}^k M = 0$. This is the case, for example, if R is Noetherian and M is finitely generated. If M is separated and the images of $m_1, \ldots, m_n \in M$ generate $M/\mathfrak{m}M$, show that $m_1, \ldots, m_n$ generate M. This is a version of Nakayama's lemma that works without assuming that M is finitely generated in advance.

Exercise 7.3: Recall that the **Jacobson radical** of a ring R is defined to be the intersection of all maximal ideals of R. Let R be a ring, and let $I \subset R$ be an ideal. Show that I is contained in the Jacobson radical of R iff all the elements of $U = \{1 + a | a \in I\}$ are invertible in R. Use this to show that if R is complete with respect to an ideal $\mathfrak{m}$, then $\mathfrak{m}$ is contained in every maximal ideal of R.

Exercise 7.4: Suppose that $R \subset S$ are Noetherian rings such that R is complete with respect to the ideal $\mathfrak{m} \subset R$, and $\mathfrak{m}$ is contained in the Jacobson radical of S. Let M be a finitely generated S-module. Show that if $M/\mathfrak{m}M$ is finitely generated as an $(R/\mathfrak{m})$-module, then M is finitely generated as an R-module. This result is most often used when $M = S$. Use it to give a different proof of the existence of the map σ in the example at the very beginning of this chapter.

Modules Whose Completions Are Isomorphic

Exercise 7.5 Reflection of isomorphism from the completion: The following result shows the tight relationship between module theory over an arbitrary local ring and over its (in many ways much better behaved) completion. Suppose M and N are finitely generated modules over a Noetherian local ring R whose completions $\hat{M}$ and $\hat{N}$ are isomorphic over $\hat{R}$. Show that $M \cong N$ as follows:

a. Deduce from Proposition 2.10 that $\operatorname{Hom}_R(M,N)^\wedge \cong \operatorname{Hom}_{\hat{R}}(\hat{M},\hat{N})$.

b. Let $\hat{P}$ be the maximal ideal of $\hat{R}$. Show that $\hat{P}\operatorname{Hom}_{\hat{R}}(\hat{M},\hat{N})$ consists of maps that take $\hat{M}$ to $\hat{P}\hat{N}$.

c. Let $\varphi \in \operatorname{Hom}_{\hat{R}}(\hat{M},\hat{N})$ be an isomorphism. Use Nakayama's lemma and part b to show that if $\varphi' \in \operatorname{Hom}_{\hat{R}}(\hat{M},\hat{N})$ differs from φ by an element of $\hat{P}\operatorname{Hom}_{\hat{R}}(\hat{M},\hat{N})$, then φ' is an epimorphism.

d. Show that there are elements $\varphi' \in \operatorname{Hom}_R(M,N)$ and $\varphi'' \in \operatorname{Hom}_R(N,M)$ that are epimorphisms. Apply Corollary 4.4a to $\varphi'\varphi''$ and to $\varphi''\varphi'$.

The Krull Topology and Cauchy Sequences

Given a descending filtration $R = \mathfrak{m}_0 \supset \mathfrak{m}_1 \supset \mathfrak{m}_2 \supset \ldots$ of an abelian group R by subgroups $\mathfrak{m}_j$, we define the **Krull topology** on R (with respect to the given filtration) by taking the subsets $\mathfrak{m}_j$ to be a base for the open neighborhoods of 0, and imposing the condition that addition should be continuous, so that the cosets $r + \mathfrak{m}_j$, with $r \in R$, form a base for the family of all open sets.

Exercise 7.6:* Show that any subgroup $\mathfrak{m}$ containing one of the $\mathfrak{m}_j$ is open.

Exercise 7.7: Show that if R is a ring and the $\mathfrak{m}_j$ are ideals, then multiplication is continuous.

A **Cauchy sequence** with respect to the Krull topology is a sequence of elements $r_i \in R$ such that for each open neighborhood U of 0 in R there is a number n with the property that for $i, i' > n$ we have $r_i - r_{i'} \in U$. Two Cauchy sequences r_i and r'_i are **equivalent** if for each open neighborhood U of 0 in R there is a number n with the property that for $i > n$ we have $r_i - r'_i \in U$.

Exercise 7.8: Show that the set of Cauchy sequences forms a group under componentwise addition, and that two sequences are equivalent iff their difference is equivalent to 0.

Exercise 7.9:* Show that the set of sequences equivalent to 0 forms a subgroup, so that the set of Cauchy sequences modulo those equivalent to 0 is again a group. We shall temporarily denote it by $\hat{R}'$. Prove that $\hat{R} \cong \hat{R}'$.

Exercise 7.10: Show that under the hypotheses of Lemma 7.14, the topologies on R defined by the filtrations $\mathfrak{m}_j$ and $\mathfrak{n}_j$ are the same. Use this to give a Cauchy sequence proof of the lemma.

Completions from Power Series

Exercise 7.11:* Let R be a Noetherian ring, and let $\mathfrak{m} = (a_1, \ldots, a_n)$ be an ideal in R. Show that

$$\hat{R}_{\mathfrak{m}} \cong R[[x_1, \ldots, x_n]]/(x_1 - a_1, \ldots, x_n - a_n).$$

Exercise 7.12: If R is Noetherian, show that any element of the form $x - a$, with $a \in R$, is a nonzerodivisor on $R[[x]]$. This is trivial if we replace $R[[x]]$ by $R[x]$, so it follows for $R[[x]]$ by the flatness of completion. Give a direct proof, without using the flatness of completion or the Artin-Rees lemma. Construct a counterexample to the statement without the Noetherian hypothesis. (Hint: The Noetherian hypothesis implies

$$(0 : a^m) = (0 : a^{m+1}) \quad \text{for large } m.)$$

Exercise 7.13: If I is a finitely generated ideal of R, show that $IR[[x]]$ is the ideal of all power series having their coefficients in I. Find an example where I is not finitely generated and the conclusion fails.

Exercise 7.14:* Taking the isomorphism of Exercise 7.11 as a definition, show directly that the completion is flat as an R-module.

Coefficient Fields

Exercise 7.15: The **characteristic** of a ring R is the positive integer that generates the kernel of the natural homomorphism $\mathbf{Z} \to R$. Let R be a local ring with residue class field K. Prove that R contains some field k, iff the characteristic of R is the same as that of K. In this case R is said to be an **equicharacteristic** local ring.

Exercise 7.16: Let $f \in \mathbf{Q}[x]$ be an irreducible polynomial of degree greater than 1 with rational coefficients, and let R be the local ring $\mathbf{Q}[x]_{(f)}$. Show that R has no coefficient field. If $f(x) = x^2 + 1$, find the (unique) coefficient field in $\hat{R}$.

Exercise 7.17 (Coefficient fields and maximal subfields): Any coefficient field in a local ring R is a maximal subfield (a subfield is a subring that is a field). By Zorn's lemma, every local ring contains maximal subfields.

a. Show that if R is complete and contains a field of characteristic 0, then every maximal subfield is a coefficient field.

b. Let $R = k(t)[[x]]$, where k is a field of characteristic $p > 0$. Show that $k(t^p + x)$ is a maximal subfield of R that is not a coefficient field. This

example shows that the hypothesis of separability cannot be deleted in Theorem 7.8.

Exercise 7.18: Suppose $X \subset Y$ is an irreducible algebraic subset of an algebraic set Y over a field k. Let R be the affine coordinate ring of X and let S be that of Y, so that $R = S/P$, for some prime ideal P. The quotient field $K := K(R)$ is the residue class field of S_P.

a. Show that the local ring of Y along X has a coefficient field containing k if, after perhaps removing a closed set from X and Y, X is a neighborhood retract in Y. To restate the problem algebraically, suppose that there is a coefficient field of S_P containing k; that is, suppose that there is a map of k-algebras $\sigma : K \to S_P$ splitting the surjection $S_P \twoheadrightarrow K$. Show there is a single element $f \notin P$ and a map $\sigma' : R \to S[f^{-1}]$ of k-algebras such that σ is the localization of σ'. (This map σ' corresponds to a map $Y - \{y \in Y | f(y) = 0\} \to X - \{x \in X | f(x) = 0\}$ that is a retraction of the inclusion map.)

b. Supposing that $S = k[x, y]$, and $P = (x^2 - y^3)$, show that there is a coefficient field, and a retraction
$$k[x, y]/(x^2 - y^3) \to k[x, y][y^{-1}].$$

c. If $K \cong k(t_1, \ldots, t_n)$, the field of rational functions in n variables, show that S has a coefficient field. Now assume that k is algebraically closed, that $S = k[x_1, \ldots, x_r]$, and that K has transcendence degree 1 over k. If you know some algebraic geometry (say Hurwitz' theorem, Hartshorne [1977, Chapter IV, Section 2]), you may show that S_p has a coefficient field iff $K \cong k(t)$.

Other Versions of Hensel's Lemma

The following theorem is the classic version of Hensel's lemma, a criterion for the factorization of a polynomial into relatively prime factors over a complete local ring.

Theorem 7.18. *Let R be a Noetherian ring, complete with respect to an ideal $\mathfrak{m}$. Let $F(x) \in R[x]$ be a polynomial in one variable with coefficients in R, and let $f(x)$ be the polynomial over $R/\mathfrak{m}$ obtained by reducing the coefficients of f mod $\mathfrak{m}$. If f factors as*
$$f = g_1 g_2 \in (R/\mathfrak{m})[x]$$
in such a way that g_1 and g_2 generate the unit ideal, and g_1 is monic, then there is a unique factorization

$$F = G_1 G_2 \in R[x]$$

such that G_1 *is monic and* G_i *reduces to* $g_i \bmod \mathfrak{m}$.

The next two exercises give a proof for this theorem.

Exercise 7.19:* Suppose that S is a ring, that $g_1, g_2 \in S[x]$ are polynomials such that g_1, g_2 together generate the unit ideal, and that g_1 is monic of degree d. Show that:

a. If $h \in S[x]$ is any polynomial, then there is a unique expression

$$h = h_1 g_1 + h_2 g_2 \quad \text{for polynomials } h_1, h_2 \quad \text{with } \deg h_2 < d.$$

(Note that the usual division with remainder is the case $g_2 = 1$.)

b. If $S = R/\mathfrak{m}$ for some ring R and ideal $\mathfrak{m}$ with $\mathfrak{m}$ in the Jacobson radical of R, and $G_1, G_2 \in R[x]$ are any polynomials such that G_i reduces mod $\mathfrak{m}$ to g_i and G_1 is monic, then G_1, G_2 together generate the unit ideal of $R[x]$.

Exercise 7.20:* Prove Theorem 7.18 by making a convergent sequence of approximate factorizations, as follows: For the first approximation, take any polynomials $G_1', G_2' \in R[x]$ with G_1' monic that reduce to g_1 and g_2 mod $\mathfrak{m}$. Show that G_1' and G_2' generate the unit ideal.

By part a of Exercise 7.19 we may write the difference between F and $G_1'G_2'$ in the form

$$F - G_1'G_2' = G_1'H_1 + G_2'H_2 \quad \text{with } H_i \in R[x], \deg H_2 < \deg G_1'.$$

As the second approximation, take $G_1'' = G_1' + H_2$ and $G_2'' = G_2' + H_1$. Show that these polynomials give a factorization of $F \bmod \mathfrak{m}^2$. Show that both H_1 and H_2 have coefficients in $\mathfrak{m}$, so that G_i'' agrees with $G_i' \bmod \mathfrak{m}$.

Since R is also complete with respect to $\mathfrak{m}^2$, we may now replace $\mathfrak{m}$ by $\mathfrak{m}^2$ in the previous argument, which results in a third approximation G_1''' and G_2''' congruent to the second mod $\mathfrak{m}^2$. Continuing this way, show that the sequence of approximations $G_i', G_i'', \ldots$ converges to polynomials G_i with the desired properties.

Exercise 7.21: Prove that Theorem 7.18 implies Theorem 7.3 in the special case where $f'(a)$ is a unit, as follows: Writing $\bar{\ }$ for reduction mod $\mathfrak{m}$, show that we may write

$$\bar{f}(x) = (x - \bar{a})g(x).$$

Now show that

$$g(x) \equiv \bar{f}'(x) \equiv \bar{f}'(\bar{a}) \bmod (x - \bar{a}).$$

Use the fact that $\bar{f}'(\bar{a})$ is a unit in $R/\mathfrak{m}$, to show that $g(x)$ and $x - \bar{a}$ generate the unit ideal in $R/\mathfrak{m}[x]$. Now lift to a factorization of $f(x)$.

Exercise 7.22 (Lifting idempotents and factorization): Here is a generalization of the case of Theorem 7.18 where F is monic, proved by lifting idempotents:

Let R be a Noetherian ring, complete with respect to an ideal $\mathfrak{m}$. Let S be an R-algebra, and $F \in S$ an element. Suppose that $A = S/(F)$ is finite as an R-algebra (that is, finitely generated as an R-module).

Let $f \in \bar{S} := R/\mathfrak{m} \otimes_R S$ be the image of F, and suppose that f factors as $f = g_1 g_2$, where $(g_1, g_2) = (1)$. Show that $\bar{S}/(f) = \bar{S}/(g_1) \times \bar{S}/(g_2)$, so that there are orthogonal idempotents $e_1, e_2 \in \bar{S}/(f)$ such that $(\bar{S}/(f))e_i = \bar{S}/(g_i)$. Lift the e_i to orthogonal idempotents E_i of $S/(F)$ so that $S/(F) = A_1 \times A_2$ with $A_i = (S/(F))E_i$.

Deduce that (with an obvious abuse of notation) $(F) = (F, E_1) \cap (F, E_2) = (F, E_1)(F, E_2)$, and that the ideals $(F, E_1), (F, E_2)$ reduce $\operatorname{mod} \mathfrak{m}$ to the ideals (g_2), (g_1), respectively. If (F, E_2) is a principal ideal, generated say by an element $G_1 \in S$, then $F \in (G_1)$, so we may write $F = G_1 G_2$ for some $G_2 \in S$. If g_1 is a nonzerodivisor, it follows that G_2 maps to $g_2 \operatorname{mod} \mathfrak{m}$, and we have thus lifted the given factorization of f.

Now suppose that S/F is free over R. Show that $S/(F, E_2)$ is projective over R. Suppose further that $\bar{S}/(g_1)$ is free over $R/\mathfrak{m}$. Show using Nakayama's lemma that $S/(F, E_2)$ is free over R of the same rank.

To deduce Theorem 7.18 when F is monic, take $S = R[x]$. By Proposition 4.1, $S/(F)$ is finite and free over R. Also by Proposition 4.1, the ideal (F, E_2) is principal and generated by a monic polynomial if $S/(F, E_2)$ is finite and free over R.

Exercise 7.23 (Direct proof of lifting idempotents):* Suppose $\mathfrak{m}^2 = 0$, and let $\bar{e} \in R/\mathfrak{m}$ be an idempotent. Find a polynomial $p(x)$ such that if $e \in R$ is any element that maps to $\bar{e}$ in $R/\mathfrak{m}$, then $p(e)$ is the unique idempotent lifting $\bar{e}$ whose existence is guaranteed by Corollary 7.5.

Exercise 7.24: Let k be a field and let R be either $k[[t]]$ or $k[t]/(t^2)$. Show that the ring of 2×2 matrices over R contains many distinct idempotents reducing mod t to the idempotent

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Exercise 7.25: Using the proof of Corollary 7.17 as a guide, prove the inverse function theorem, which says:
Let $f_1, \ldots, f_n \in (x_1, \ldots, x_n)R[[x_1, \ldots, x_n]]$ be n power series. If φ is the endomorphism

$$\varphi : R[[x_1, \ldots, x_n]] \to R[[x_1, \ldots, x_n]], \quad x_1, \ldots, x_n \mapsto f_1, \ldots, f_n,$$

and $J(x)$ is the Jacobian matrix

$$J(x) = (\partial f_i / \partial x_j),$$

then φ is an isomorphism iff det $J(0)$ is a unit in R.

Exercise 7.26: With notation as in Exercise 7.25, suppose det $J(a)$ is a unit in R. Follow the outline of the proof of Theorem 7.3 to show that if $(a_1, \ldots, a_n) \in R^n$ is an approximate solution to the system of equations $f_i(x) = 0$ in the sense that

$$f_i(a_1, \ldots, a_n) \in \mathfrak{m} \quad \text{for } i = 1 \text{ to } n,$$

then there is an actual solution $(b_1, \ldots, b_n) \in R^n$ of the equations such that each b_i differs from a_i by an element of $\mathfrak{m}$. If you feel ambitious, do the same without assuming that det $J(a)$ is a unit, but assuming instead that each $f_i(a_1, \ldots, a_n)$ is in the ideal

$$(\det J(a))^2\mathfrak{m}.$$

See Bourbaki [1985, sect 4.6, Theorem 2] if you get stuck. (Most of the above treatment is taken from this source.)

Exercise 7.27:* Give a criterion for a p-adic unit u to be an nth power for any n.

Part II

Dimension Theory

In Chapter 8 I have given an introduction to dimension theory from an historical point of view, and I have tried to explain the geometry behind some of the most important results and definitions. This chapter could logically be skipped, or postponed until later.

Chapters 9 through 14 present a unified treatment of classical dimension theory. By contrast, Chapter 15, Gröbner Bases, is a general introduction to that subject, which could be read independently of the preceding chapters; indeed, it may be read independently of the rest of this book. I have included it in this part because the technique it contains allows one to compute dimensions explicitly (it is not obvious from the definition of dimension that effective computation is possible at all). Chapter 16 presents the technique of differentials. Related to the tangent bundle, this is another fundamental technique for handling dimension.

8
Introduction to Dimension Theory

> Of all the theorems of analysis situs, the most important is that which we express by saying that space has three dimensions. It is this proposition that we are about to consider, and we shall put the question in these terms: When we say that space has three dimensions, what do we mean?
>
> —Henri Poincaré, quoted by Hurewicz and Wallman [1941]

As with Chapter 1, the material presented in this chapter is rather advanced compared to the rest of this book. If you have never studied dimension theory before, you may find it difficult to understand the material in detail. I suggest that you browse through Chapter 8 without worrying about the details during the first reading; I hope that it will tell you something of what is significant in the theory. In Chapter 9 I have begun the subject again, with a self-contained and more elementary account. None of the actual results and definitions in Chapter 8 will be required for understanding the rest of the book.

Arguably the most fundamental notion in geometry and topology is that of dimension. In this part of the book we shall take up its algebraic analogue, which plays an equally fundamental role in commutative algebra and algebraic geometry. In this section we shall sketch a little of the history that led to the modern algebraic notion, called Krull dimension, and explain some of the reasons for accepting it as the "right" definition, at least for Noetherian rings. (This explanation, beginning with the Axioms D1–D4 leads us into rather advanced territory, and will not be used in the sequel.) We then outline some of the central results of the theory. In the

following chapters we shall start again and give a self-contained and more elementary account.

To help understand the differences as well as the similarities between the algebraic and topological notions, we begin by discussing the topological one. It was an idea of the ancient Greek mathematicians that a curve (we would say a curve segment) was something bounded by points; that a surface was something bounded by curves; and that a volume was something bounded by surfaces. In nineteenth-century geometry, the idea of dimension was used intuitively. Euclidean n-space was, by agreement, of dimension n; and in general an object was said to be n-dimensional if the least number of parameters needed to describe its points, in some unspecified way, was n. Knowledge of set theory, necessitated by the increasing sophistication of analysis, caused the geometers to be driven from this paradise near the end of the century: Cantor's one-to-one correspondence between the points of a line and the points of a plane (1875), and the space-filling curves of Peano (1890) and Hilbert, showed decisively that more subtle ideas were necessary. These developments must have been quite unsettling: Cantor himself wrote of the one-to-one correspondence in a letter to Dedekind in 1877:

> Your latest reply about our work was so unexpected and so novel that in a manner of speaking I will not be able to attain a certain composure until I have had from you, my very dear friend, a decision on its validity. As long as you have not confirmed it, I can only say: *I see it but I don't believe it.* [$\cdots$] the distinction between domains of *different* dimensions must be sought for in quite another way than by the characteristic number of independent coordinates. (The translation is from Fauvel and Gray [1987], or see Purkert and Ilgauds [1935] pp. 32–35.)[1]

A precise topological definition of dimension was first given by L.E.J. Brouwer (1913), working from ideas of Poincaré. In 1922, Menger and Urysohn independently found a similar definition, which coincided with Brouwer's for most spaces: Dimension is a local property of a space at a point and is defined inductively to be the smallest number n for which arbitrarily small neighborhoods of the point have boundary of dimension less than n, To start things off, the empty set is defined to have dimension -1. (A beautiful exposition of the topological theory may be found in the classic Hurewicz and Wallman [1941], from which the quotation at the head of this chapter is borrowed.)

In algebraic geometry the notion of dimension has some special peculiarities. From the study of the conic sections in antiquity until about 1800,

---

[1]Reprinted from *The History of Mathematics: A Reader* by J. Fauvel, J. Gray (1987), by permission of MacMillan Press Ltd., London, England.

algebraic geometry concerned itself with real algebraic curves. After the introduction of coordinates by Descartes, such curves were defined by one equation on the coordinates of the two-dimensional Euclidean plane and had dimension 1 in every sense. (Since the curve rather than the equation was fundamental, the fact that some equations, like $x^2 + y^2 = 0$, do not define a curve in the real plane was unimportant.) The introduction of complex numbers and the complex projective plane in the first third of the nineteenth century changed the nature of the objects considered, but not the view that they were one-dimensional. Curves now had complex points (for example, a circle $r = x_1^2 + x_2^2$, thought of as a subset of the projective plane with equation $rx_0^2 = x_1^2 + x_2^2$, contained the famous "circular points at infinity" $(0, 1, \pm i)$ in the complex projective plane—see Exercise 1.15c). Thus the idea was born that dimension had a meaning independent of what field is used for the coordinates of points.

As far as I am aware, the early workers were not concerned with the collection of all complex points of a plane curve as a single geometric object, a surface. But that concern took the spotlight in the work of Riemann, as understood by Clebsch [1864] and later authors, where algebraic curves, interpreted as "Riemann surfaces," arose as coverings of $\mathbf{C} = \mathbf{R}^2$ or better of the complex projective line, or "Riemann sphere" $\mathbf{C} \cup \{\infty\}$. Now $\mathbf{C}$ could reasonably be called either the *complex plane* (from a topological point of view) or the *complex line* (from a complex-analytic point of view). Thus algebraic curves had reasonable claims to being called either one- or two-dimensional. Even the names, Riemann **surfaces** and algebraic **curves**, suggest a certain schizophrenia. We have grown used to the confusion, and speak happily in general of an n-dimensional complex manifold, which, as a topological space, has dimension $2n$.

In the study of a Riemann surface X, the field $K(X)$ of meromorphic functions on X was important from the start, as was the result that this field has transcendence degree 1 as an extension of $\mathbf{C}$. The field $K(X)$ coincides with what we would call the field of rational functions on X; if X is given as an affine plane curve with equation $f(x, y) = 0$, then $K(X)$ is the quotient field of the domain $\mathbf{C}[x, y]/(f)$.

In the last third of the nineteenth century, a good deal of attention was also given to spaces Y described as the *zero loci* of single equations in complex three-space; these were considered to be surfaces by algebraic geometers, although they are four-dimensional in the topological sense. Again, the field of rational functions $K(Y)$ took center stage. The fact that $K(Y)$ has transcendence degree 2 over $\mathbf{C}$ was interpreted by saying that it takes two complex-valued algebraic functions to parametrize the points of Y (up to finite ambiguity), and thus that Y has dimension 2.

From these beginnings, and from the axiom, like that of the topologists, that affine d-space has dimension d, came the algebraic definition of dimension that was used early in this century: The dimension of an irreducible variety in affine r-space over a field k (initially $\mathbf{C}$) is the transcendence

degree over k of the field of rational functions on X. This is by definition the quotient field of the domain $R = k[x_1, \ldots, x_r]/I$, where I is the ideal of all functions vanishing on X. It was only natural to define the dimension of R to be this same transcendence degree. This definition was still the accepted standard as late as 1935, as one may see from Krull's famous book *Idealtheorie,* published in that year.

The definition of dimension as transcendence degree over a ground field k agrees with our modern notion as long as one sticks to the coordinate rings of affine algebraic varieties over k: that is, to domains finitely generated over k. However, it is inadequate for other fundamental examples. For instance, rings of algebraic numbers (finite extensions of the ring of integers) do not even contain fields. Also, geometric examples involving reducible algebraic sets, such as that of Figure 8.1 make it clear that dimension is most interesting as a local property of a space at a point. Unfortunately, the localization of the coordinate ring of a variety, which might be hoped to be the carrier of this local information, is almost never finitely generated over the ground field. If one passes to power series rings, which represent the variety in a local analytic sense, the situation is still worse; $k[[x]]$ has uncountable transcendence degree over k, although its dimension should, on geometric grounds, be 1.

Although the definition by transcendence degree is useless in the case of rings of algebraic numbers, there is an analogy that suggests what the dimensions of these rings should be. As we saw in Chapter 1, some of the very earliest algebraic work on Riemann surfaces was done by exploiting the amazing analogy between Riemann surfaces and rings of algebraic numbers. On the basis of this analogy one might well imagine that the dimension of any ring of algebraic numbers should be defined to be 1. In other cases where the definition by transcendence degree is not the right thing, other ad-hoc arguments could be made. For the local ring of a point on an algebraic variety, for example, the definition of the dimension can be taken from that of the variety itself. The same reasoning could be made

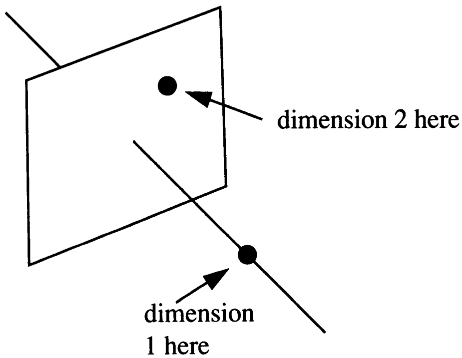


FIGURE 8.1.

for power series rings and their factor rings, which correspond to points on analytic varieties.

As increasingly complex constructions were made in commutative algebra, such ad-hoc definitions became unsatisfactory. In 1937 Krull proposed the following definition. For justification he quoted geometric evidence accumulated by Emmy Noether [1923] for factor rings of polynomial rings and by W. Rückert [1932] for factor rings of power series rings. He also mentioned the arithmetic analogy given in the previous paragraph. The definition now bears his name.

Definition. *The* ***Krull dimension*** (*or simply the* ***dimension***) *of a ring R, written* **dim *R***, *is the supremum of the lengths of chains of distinct prime ideals in R.*

Here the length of the chain $P_r \supset P_{r-1} \supset \cdots \supset P_0$ involving $r+1$ distinct prime ideals is taken to be r; we remind the reader that the ring itself is not considered a prime ideal. The supremum may in fact be infinity even for Noetherian rings (see Exercise 9.6), although in the case of a Noetherian local ring or an affine ring we shall see that it is finite.

This definition gains plausibility from the familiar fact that the dimension of a vector space over a field k is the length of the longest chain of proper subspaces; algebraically, an n-dimensional vector space corresponds to a polynomial ring $R = k[x_1, \ldots, x_r]$, and an increasing sequence of subspaces, starting with 0, corresponds to the decreasing sequence of prime ideals

$$(x_1, \ldots, x_r) \supset (x_1, \ldots, x_{r-1}) \supset \cdots \supset (x_1) \supset 0.$$

We shall see in Theorem A that no chain of greater length exists.

As first examples, we see that any field k has dimension 0, while the polynomial ring $k[x]$ has dimension 1: $(x) \supset 0$ is a chain of primes of length 1, and since every ideal is principal, there are no longer chains of primes. (*Reason:* If $(p) \supseteq (q) \supsetneq 0$ with (q) a prime ideal, then q is a prime element. Since p divides q, the elements p and q differ by a unit, whence $(p) = (q)$.) The same argument applies in any principal ideal domain that is not a field, so, for example, the ring $\mathbf{Z}$ of integers has dimension 1.

It is interesting to compare Krull's definition with Menger's, given earlier, although this is unhistorical in the sense that Krull does not mention the topological case or its pioneers. Krull's definition exploits the simple structure of algebraic varieties and replaces the closed sets that are the boundaries of small neighborhoods of p that occur in Menger's definition by maximal closed sets not containing any component *and* containing p. If we regard the set of primes of a ring as a topological space in the Zariski topology and apply Menger's definition, we get the same thing in many, but not all, cases; Krull's definition is in a certain sense more local.

Is Krull's definition the "right" definition of dimension? We offer two kinds of responses to this vague question:

First, we shall give a few axioms satisfied by the definition, corresponding to simple geometric properties. The axioms determine the notion of dimension uniquely. (The proof of this fact depends on the Cohen structure theorems, and we shall only sketch it.) We give them in order to orient the reader toward the central properties of dimension—especially its behavior under ring extensions, of which Axiom D3 is the simplest part. The axioms themselves will not be used in the remainder of this book.

Second, we shall list some characterizations of dimension that serve to connect the notion with other geometric ideas and make the concept fruitful. These characterizations play a major role in the rest of this book, and some of the succeeding sections will be organized around proofs of them.

We shall assume for the rest of this section, and in nearly all of Part II, that all rings considered are Noetherian. In the non-Noetherian case the Krull dimension exhibits various pathologies. For example, if R is Noetherian, then $\dim R[x] = (\dim R) + 1$ (Corollary 10.13). Geometrically, if R is the ring of functions on some space, then $R[x]$ is the ring of functions on the product of that space and the affine line, so this formula is forced. But in the non-Noetherian case one can have $\dim R[x] = (\dim R) + 2$. See, for example, Gilmer [1974] for a study of this phenomenon.

8.1 Axioms for Dimension

1. Just as in the topological definition of dimension given above, dimension should be a local property. This means, in particular, that the dimension of a ring should be the maximum of the dimensions of its localizations; but also, since passing to the completion corresponds geometrically to taking a smaller neighborhood, that dimension at a point is preserved by completion.

Axiom D1 (Dimension is a local property).

$$\dim R = \sup_{P \textit{ is a prime of } R} \dim R_P,$$

and

$$\dim R_P = \dim \hat{R}_P.$$

2. If X is an affine algebraic set then the ring of functions on X has no nilpotent elements; nilpotent functions are not apparent in ordinary geometry. On the other hand, if R is the affine ring of an algebraic set X, and $p \in X$ corresponds to the maximal ideal P in R, so that $R/P = k$ is the "ring of functions on $\{p\}$," then it is possible to regard an element $f \in R/P^2$ as a function with a value at p together with a linear functional on the tangent space to X at p. Thus as a ring, R/P^2 describes first-order jets of functions on X at p (this will be explained further in Chapter 16). We may think of it as the ring of functions on a first-order

infinitesimal neighborhood of p. In the theory of schemes (see, for example, Eisenbud and Harris [1992]) nilpotent elements of an affine ring R are quite generally interpreted as describing some infinitesimal neighborhood of the variety defined by $R_{\text{red}} = R/(\text{nilpotent elements})$. Thus it is geometrically reasonable to pose as an axiom:

Axiom D2 (Nilpotents do not affect dimension). *If I is a nilpotent ideal of R, then* $\dim R = \dim R/I$.

3. For a $\mathcal{C}^\infty$ surjective map $f : M \to N$ of manifolds, Sard's theorem implies that the general fiber of f (fiber $\equiv$ preimage of a point) is again a manifold, and the dimension of M is the dimension of N plus the dimension of the general fiber. Such a principle also holds, under mild assumptions, for the Krull dimension, but all we need here is the case corresponding to a surjective map with finite fibers (in which the dimension of M will equal the dimension of N). To see what the algebraic content of this condition is, consider a map

$$\varphi : R = k[x_1, \ldots, x_r]/J \to k[x_1, \ldots, x_s]/J' = S$$

corresponding to a map $\psi : Y \to X$ of algebraic varieties. The image of ψ is dense in X (in the Zariski topology; but if $k = \mathbf{C}$, also in the classical topology) iff φ is a monomorphism. We shall show that if φ makes S a finitely generated R-module, then the fibers of ψ are all finite sets (Corollary 9.3). Under these circumstances, we shall have $\dim X = \dim Y$. We take such behavior as the third axiom.

Axiom D3 (Dimension is preserved by a map with finite fibers). *If $R \subset S$ are rings such that S is a finitely generated R-module, then* $\dim R = \dim S$.

4. Finally, in examples where we have some reason to know what the dimension should be, we should get the expected answer. First, since the polynomial ring $k[x_1, \ldots, x_r]$ in r variables over a field k corresponds to affine r-space, it should have dimension r; as the formal power series ring $k[[x_1, \ldots, x_r]]$ is its completion at the maximal ideal $(x_1, \ldots, x_r)$, this ring should have dimension r. The converse implication works too: If k is algebraically closed, then a consequence of the Nullstellensatz (Exercise 4.28) shows that any maximal ideal of $k[x_1, \ldots, x_r]$ can be transformed by an automorphism into $(x_1, \ldots, x_r)$, and thus $k[x_1, \ldots, x_r]$ has dimension r if $k[[x_1, \ldots, x_r]]$ does. Thus, to cover both examples, we need only make an assumption about one. To emphasize that dimension is a local property, we choose the power series ring.

Axiom D4a (Calibration—algebras over a field). *If k is a field, then* $\dim k[[x_1, \ldots, x_r]] = r$.

This suffices if we only want to work with rings containing a field. To include the arithmetic case we need a slightly more sophisticated version, one that includes the idea that rings of algebraic integers have dimension 1. Since we have taken dimension to be a local property, we only need to assert this for the localizations, or even the completions, of such rings. These have the property that they are local Noetherian integral domains in which every ideal is principal, but not fields; such rings are called **discrete valuation rings**. A complete discrete valuation ring that contains a field is isomorphic to a formal power series ring in one variable over a field (see Proposition 10.16), and thus is certainly of dimension 1. A part of the Cohen structure theorems that we do not treat in this book shows that all other examples look a lot like the ring $\hat{\mathbf{Z}}_p$ of p-adic integers, or a finite extension of such a ring. Now since

$$k[[x_1, \ldots, x_r]] = k[[x_1]][[x_2, \ldots, x_r]],$$

we may think of a power series ring in r variables as a power series ring in $r-1$ variables over a discrete valuation ring, and if we take the position that all discrete valuation rings should have analogous dimension-theoretic properties then we arrive at the following stronger version of Axiom D4a.

Axiom D4b (Calibration—general case). *If R is a complete discrete valuation ring, then*

$$\dim R[[x_2, \ldots, x_r]] = r.$$

These properties suffice to characterize a function "dim" on the class of Noetherian rings, and this function is equal to the Krull dimension. Since we shall not return to these axioms, it seems worthwhile to sketch the proof (which may only be intelligible after this book has been read): By Axiom D1 and the fact that the completion of a Noetherian ring is Noetherian (see Theorem 7.1), we need only consider complete local Noetherian rings. By Axiom D2 we may assume that they have no nilpotent elements. But as we shall see (Exercise 13.9), a complete Noetherian ring containing a field is a finite module over some subring isomorphic to a power series ring in finitely many variables over a field. Its dimension is thus determined by Axioms D3 and D4a; a similar result that we have not proved (references are given in Chapter 7) shows that a complete local ring without nilpotent elements that does not contain a field is a finitely generated module over a subring isomorphic to a power series ring over a complete discrete valuation ring. Its dimension is thus determined by Axioms D3 and D4b.

8.2 Other Characterizations of Dimension

At the heart of dimension theory are three further characterizations of dimension that are useful in different contexts. We shall return to these in later chapters.

8.2.1 Affine Rings and Noether Normalization

We may characterize dimension for affine rings as follows:

Theorem A. *If R is an affine domain over a field k, then*

$$\dim R = \textit{transcendence degree}_k R.$$

This is the common length of all maximal chains of prime ideals of R.

In particular, this shows that the dimension of an affine domain is finite.

The first part of this statement is an avatar of the idea that the dimension of a variety is the number of independent functions on it. In this sense the algebraic geometers have never left paradise: There is no snake (that is, Peano curve) in the garden. The reason is that algebraic geometers work with such a restricted class of functions. The second part is a uniformity result, a strengthening of the idea that the dimension of an irreducible variety is the same at each of its points. It implies for example that if an affine ring contains a chain of four primes $P \supsetneqq P_1 \supsetneqq P_2 \supsetneqq Q$ and a prime P' between P and Q as in Figure 8.2, then it contains either a prime strictly between P and P' or between P' and Q. In technical language we say that an affine ring is **catenary**. Not every Noetherian ring is catenary; this is one of the important "nongeometric" pathologies that Noetherian rings exhibit (see Nagata [1962] for an example). But the catenary condition holds for virtually any ring that one could meet in algebraic geometry. We shall prove a general result along these lines in Corollary 18.10.

We shall prove Theorem A using the Noether normalization theorem (13.3), which says that any affine ring of dimension r and any chain of primes in it are comparable, via a finite map, to a polynomial ring $k[x_1, \ldots, x_r]$ and a chain whose members are primes generated by subsets of the variables. More precisely,

Theorem A1 (Noether Normalization). *If R is an affine ring over a field k (that is, R is a finitely generated k-algebra) and $P_r \supsetneqq P_{r-1} \supsetneqq \cdots \supsetneqq P_0$ is a chain of prime ideals of R, maximal in the sense that no further prime*

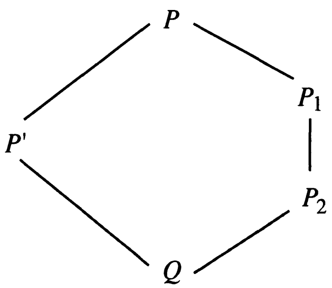


FIGURE 8.2.

ideals can be inserted into the chain, then there is a subring S of R with $S \cong k[x_1, \ldots, x_r]$ such that R is a finitely generated S-module and $P_i \cap S = (x_1, \ldots, x_i)$.

The Noether normalization theorem is a touchstone for results on affine rings. We apply it to give another proof of Hilbert's Nullstellensatz, to prove the finiteness of the integral closure of an affine domain, and to prove some results on the behavior of the fibers of a map.

8.2.2 Systems of Parameters and Krull's Principal Ideal Theorem

Krull's principal ideal theorem (Krull [1928]) may be expressed as a characterization of dimension:

Theorem B. *If R is a Noetherian local ring with maximal ideal $\mathfrak{m}$, then $\dim R$ is the minimal number n such that there exist n elements $f_1, \ldots, f_n \in \mathfrak{m}$ not all contained in any prime other than $\mathfrak{m}$.*

To understand the geometric content of this result, consider first the case where $\mathfrak{m}$ is a maximal ideal of an affine ring S, corresponding to a point p on some algebraic variety M, and $R = S_{\mathfrak{m}}$. Considering the f_i as functions on M defined near p, the condition that no smaller prime contains all the f_i becomes the condition that the point p is singled out (in some small neighborhood of p) by the vanishing of all the f_i. Parallel in a certain sense to the characterization of dimension by transcendence degree, Theorem B is a second avatar of the idea that the dimension of R is "the least number of parameters needed to describe the points of M."

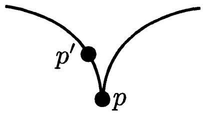
Among the many algebraic corollaries of Theorem B, one of the most striking is that there are no infinite descending chains of primes in a Noetherian ring. In fact, the number of generators of a prime ideal P gives an a priori bound for the lengths of chains of primes descending from P. In particular, this shows that if R, $\mathfrak{m}$ is a local ring then

$$\dim R \leq \text{the number of generators of } \mathfrak{m} < \infty.$$

Another appreciation of Theorem B may be had from the (leading) special case where P is minimal among primes containing a principal ideal $(f) \neq 0$ in a ring R. Localizing at P, the result says that $\dim R_P \leq 1$ (and if f is a nonzerodivisor the dimension is exactly 1). Thus, there is no chain of primes of length greater than 1 descending from P. This is the original form of the principal ideal theorem from which the name derives.

To understand the geometric content in this case, suppose that R is an affine domain over $\mathbf{C}$, corresponding to a variety M that is a complex manifold. Suppose that N is the subvariety of points x such that $f(x) = 0$.

By the second statement of Theorem A, the statement $\dim R_P = 1$ shows that $\dim M - \dim N$ is equal to 1. To prove this "by geometry," we may pick a point p on N and claim that there is a point p' near p on N (see the following figure) where the derivatives of f do not all simultaneously vanish. At p', the implicit function theorem asserts that N looks analytically like a coordinate hyperplane—and is thus codimension 1.



By Theorem A an affine variety has the same dimension at each of its points, so the dimension is 1 at p too. (To prove the claim, we may as well pass to a local analytic neighborhood of p and assume that R is a power series ring. We may further replace f by one of its factors and assume that f is irreducible in this power series ring. Of course, not all the derivatives of f can vanish identically, or f would be a constant. If $g \neq 0$ is the derivative of f in some direction, then g is a power series whose initial term is of lower degree than that of f—so g is not in the ideal generated by f. Since (f) was supposed to be prime, this and an analytic version of the Nullstellensatz show that $g(p') \neq 0$ for some p' near p in N, as claimed.)

8.2.3 The Degree of the Hilbert Polynomial

The characterizations of dimension given in Theorems A and B are both of the form that the dimension is the maximum number of functions with a certain independence property, or the minimal number with a certain sufficiency property. The following characterization avoids individual functions altogether.

Theorem C. *Let R be a Noetherian local ring with maximal ideal $\mathfrak{m}$, and let $H(n)$ be the Hilbert function*

$$H(n) = \dim_{R/\mathfrak{m}} \mathfrak{m}^n/\mathfrak{m}^{n+1}.$$

For large n, $H(n)$ agrees with a polynomial $P(n)$, and

$$\dim R = 1 + \textit{degree } P.$$

Here $\dim_{R/\mathfrak{m}}$ denotes the ordinary vector space dimension over the field $R/\mathfrak{m}$. This result shows that the dimension of R depends only on the associated graded ring of R (geometrically, we recall from Chapter 5 that if R is the local ring of a point p on a variety M, then the associated graded ring $\operatorname{gr}_\mathfrak{m} R$ corresponds to the tangent cone of M at p). Since $\operatorname{gr}_\mathfrak{m} R = \operatorname{gr}_\mathfrak{m} \hat{R}$,

where $\hat{R}$ is the completion of R at $\mathfrak{m}$, Theorem C implies that the dimension of R is the same as that of $\hat{R}$, as demanded by Axiom D1.

Geometrically, Theorem C is loosely an analogue of the fact that the dimension of a variety near a point measures how fast the volume of a neighborhood of the point grows with the diameter of the neighborhood. One way to formulate this result, due to Thie [1967] is as follows. Suppose p is a point on a complex analytic variety $M \subset \mathbf{C}^r$, and B_ε is the ε-ball around p in $\mathbf{C}^r$. The dimension of M is the unique integer d such that if V_ε is the integral over the smooth points of M in $M \cap B_\varepsilon$ of the d-dimensional volume form on $\mathbf{C}^r$, then $\lim_{\varepsilon\to 0} V_\varepsilon/\varepsilon^d$ is a finite nonzero number. (In fact, if W_ε is the volume of $\mathbf{C}^d \cap B_\varepsilon$, then $\lim_{\varepsilon\to 0} V_\varepsilon/W_\varepsilon$ is the multiplicity of M at p; see Chapter 12 for the definition.) If M is algebraic, R is the local ring of p on M, and $\mathfrak{m}$ is the maximal ideal of R, then the ring $R/\mathfrak{m}^n$ should be seen as the coordinate ring of the "nth-order infinitesimal neighborhood" of p in M, and the number $\dim_{\mathbf{C}} R/\mathfrak{m}^n = \sum_{i=1}^n H(n)$ is a measure of the size of this neighborhood—some sort of infinitesimal volume. Theorem C implies that

$$\left(\sum_{i=1}^{n} H(n)\right) \Big/ n^d$$

has a finite nonzero limit as $n \to \infty$ iff $d = \dim M$.

Though one might suspect the opposite, Theorem C actually opens the best avenue to the algorithmic computation of the dimension of an affine ring: One first homogenizes the equations defining the ring to reduce to the graded case, say $R = k[x_1, \ldots, x_r]/I$, with I homogeneous. The theory of Gröbner bases then shows how to construct a monomial ideal I' such that $k[x_1, \ldots, x_r]/I$ and $k[x_1, \ldots, x_r]/I'$ have the same Hilbert function, and thus the same dimension. Simple combinatorial ideas then suffice to compute the dimension for I'. For all of this, see Chapter 15.

9

Fundamental Definitions of Dimension Theory

In this chapter we collect the fundamental definitions and notation we shall use. We also harvest the statements on dimension theory that have been proved earlier in this book, before we had the language to describe them: the characterization of dimension zero from Chapter 2 and the properties of integral maps (relative dimension zero) from Chapter 4. To make this chapter and what follows independent of the introductory Chapter 8, we repeat a few definitions.

Definition. *The **Krull-dimension** (or simply the **dimension**) of a ring R, written* **dim *R***, *is the supremum of the lengths of chains of prime ideals in R.*

Here the length of the chain $P_r \supsetneq P_{r-1} \supsetneq \cdots \supsetneq P_0$ of prime ideals is taken to be r; we recall that the ring itself is not considered a prime ideal.

Now let R be a ring, and $I \subsetneq R$ an ideal. We define the **dimension of *I***, written **dim *I***, to be $\dim R/I$. The name corresponds to the fact that if R is the ring of functions on an algebraic set, then $\dim I$ is the dimension of subset corresponding to I; that is, the subset on which the "functions" in I vanish.

If I is prime then the **codimension of *I***, written **codim *I*** (also called height I and rank I by various authors), is by definition the dimension of the local ring R_I. Equivalently, it is the supremum of lengths of chains of primes descending from I. If I is not assumed prime, then we define $\operatorname{codim} I$ to be the minimum of the codimensions of the primes containing I. (Our terminology follows that of Krull, who called this the *Dimensionsdefekt.* He

remarks that he chose this name in place of the name "rank," which had already been used in the case of polynomial rings by Lasker and Macaulay, to emphasize the geometric content of the idea.)

If N is any R-module, then we define the dimension and codimension of N to the be the dimension and codimension, respectively, of the annihilator of N. Unfortunately, if N is an ideal, this standard definition conflicts with the equally standard definition of the dimension of an ideal I as the dimension of the ring R/I. For example, if R is a domain, then the annihilator of any nonzero ideal is trivial, so the dimension of the ideal *as a module* is equal to the dimension of R. Perhaps because these two definitions give such different answers, their simultaneous use does not seem to cause confusion. When we write $\dim I$, for an ideal I, we shall always mean the dimension of the ring R/I.

If R is a domain, finitely generated over a field, and $I \subset R$ is an ideal, then from the definitions, with Theorem A described in Chapter 8, it follows that $\operatorname{codim} I = \dim R - \dim I$. But we have not used this as the definition, because when R is not a domain, the quantity $\dim R - \dim I$ is not local in the sense that we want dimension theory to be local. To see this, consider the example portrayed in Figure 9.1 of a plane M and a line L containing a point N. Algebraically, we may represent this by

$$\begin{aligned} R &= k[X_1, X_2, X_3]/(X_1)(X_2, X_3) &&\leftrightarrow M \cup L, \\ I &= (X_1 + 1, X_2, X_3) &&\leftrightarrow N. \end{aligned}$$

Here $\dim M = \dim R = 2$, whereas $\operatorname{codim} I = 1$ and $\dim N = \dim R/I = 0$, so

$$\dim I + \operatorname{codim} I \neq \dim R.$$

Instead, $\operatorname{codim} I$ is the codimension of N in the component of M in which N lies.

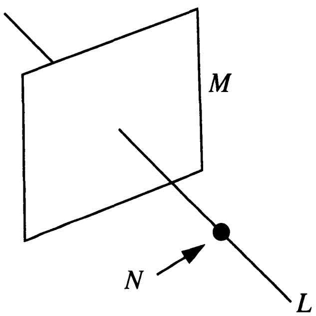


FIGURE 9.1.

9.1 Dimension Zero

We now interpret some of the earlier results of this book in terms of dimension. Theorem 2.14 and Corollary 2.15 characterize rings of dimension 0.

Corollary 9.1. *If R is Noetherian, then* $\dim R = 0$ *if and only if R is Artinian, in which case R is a direct product of local Artinian rings. An affine algebraic set has dimension 0 iff it is a finite set.*

We shall analyze affine rings by exhibiting them as finitely generated modules over polynomial rings. For this reason we need a "relative" version of Corollary 9.1. We need to understand the behavior of prime ideals with respect to a homomorphism of rings $\varphi : R \to S$ that makes S a finitely generated R-module. It is technically convenient to work with the more general condition that S is integral over R.

Proposition 9.2. *If $\psi : R \to S$ is a map of rings that makes S integral over R, then every prime ideal of R containing $\ker \psi$ is the preimage of a prime ideal of S. Furthermore, if I is an ideal of S, then*

$$\dim I = \dim \psi^{-1} I.$$

Proof. We may replace R by its image in S, and thus assume that $R \subset S$, so that we can apply Proposition 4.15 and Corollary 4.18. The first statement is immediate from Proposition 4.15.

For the dimension equality, consider a chain of primes of R ascending from $\psi^{-1}I$. By going up (Proposition 4.15), there is a chain of primes of S ascending from I and having as preimage the given chain of primes of R. Thus $\dim I \geq \dim \psi^{-1} I$.

On the other hand, by incomparability (Corollary 4.18), a chain of distinct primes containing I has as preimage a chain of distinct primes containing $\psi^{-1}I$, and thus $\dim I \leq \dim \psi^{-1} I$, whence the equality. □

Here is the geometric version.

Corollary 9.3. *If $\varphi : X \to Y$ is a morphism of affine algebraic sets such that $A(X)$ is a finitely generated $A(Y)$-module, then:*

1. *The fibers of φ are finite sets.*

2. *If $X' \subset X$ is a Zariski closed subset, then $\varphi(X') \subset Y$ is a Zariski closed subset with the same dimension as X'. In particular, if $A(Y) \subset A(X)$, then φ is surjective.*

Maps φ satisfying the hypothesis of the corollary are called **finite** maps.

Proof. Replacing X by X' and Y by the closure of $\varphi(X')$, we may assume $A(Y) \subset A(X)$ and we must show that φ is surjective and that X and Y have the same dimension. These things are immediate from Proposition 9.2. Furthermore, we see from Proposition 9.2 that φ has zero-dimensional fibers. By Corollary 9.1, zero-dimensional algebraic sets are finite. $\square$

The same proof would prove the corollary with the weaker hypothesis that $A(X)$ is integral over $A(Y)$, but since $A(X)$ is finitely generated as a ring, this makes no difference here. However, it would also prove the same statement for the map $\operatorname{Spec} S \to \operatorname{Spec} R$ for any rings R and S as in Proposition 9.2. In the context of schemes, the result is useful in this more general form.

9.2 Exercises

Exercise 9.1: Show that a principal ideal ring (one whose ideals are each generated by ≤ 1 element) has dimension ≤ 1.

Exercise 9.2: Let k be a field.

a. Let $f(x,y) \in k[x,y]$ be any polynomial, and consider the "variable" $x' = x - y^n$. Show that $k[x,y] = k[x',y]$, and that if n is sufficiently large, then as a polynomial in x' and y, f is monic in y. Deduce that $k[x,y]/f$ is integral over its subring $k[x']$. Use this to prove that $\dim k[x,y] = 2$.

b. Show that the same things are true for $x' = x - ay$ for all but finitely many $a \in k$. (If k is finite, this could be all $a \in k$.)

Exercise 9.3: Suppose that a ring S is integral over the image of a ring homomorphism $R \to S$. Show that the Krull dimension of M as an S-module is the same as the Krull dimension of M as an R-module.

Exercise 9.4: Suppose that U is a multiplicatively closed subset of R, and let $S = R[U^{-1}]$ be the localization. Let P be a prime of S. Show that $\operatorname{codim} P = \operatorname{codim}(P \cap R)$.

Exercise 9.5 (Dimension of Veronese subrings): Let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded Noetherian ring, and let $R_{(d)} = R_0 \oplus R_d \oplus R_{2d} \oplus \cdots$, the **Veronese subring**. Show that R is integral over $R_{(d)}$; conclude that $\dim R = \dim R_{(d)}$. Show that there is a one-to-one correspondence between the homogeneous primes of R and the homogeneous primes of $R_{(d)}$ defined by $R \supset P \mapsto P \cap R_{(d)}$. Also, show that if $x \in R$ is a homogeneous element of strictly positive degree, then $R[x^{-1}]_{(d)} = R_{(d)}[x^{-d}]$; thus, in particular,

$(R[x^{-1}])_0$, the degree 0 part, is equal to $(R_{(d)}[x^{-d}])_0$. (For geometers: This says that $\text{Proj}(R) = \text{Proj}(R_{(d)})$ as schemes.) Taking $R = k[x]$, where k is a field, show by example that the correspondence $P \mapsto P \cap R_{(d)}$ may be many-to-one for nonhomogeneous prime ideals.

Exercise 9.6 (An infinite-dimensional ring):* One of the pathologies that Noetherian rings can have is that they can be infinite-dimensional (although as we shall later prove, a Noetherian local ring must be finite-dimensional). Here is an example due to Nagata [1962] (Appendix, example E1): Let $R = k[x_1, \dots, x_r, \dots]$ be a polynomial ring in infinitely many variables over a field k, and let $P_1 = (x_1, \dots, x_{d(1)}), P_2 = (x_{d(1)+1}, \dots, x_{d(2)}), \dots, P_m = (x_{d(m-1)+1}, \dots, x_{d(m)}), \dots$ be an infinite collection of prime ideals made from disjoint subsets of the variables. Let $U = R - \cup_{m=1}^{\infty} P_m$ be the complement of the union of the primes P_m, and let $S = R[U^{-1}]$. By Exercise 3.18 the maximal ideals of S are precisely the ideals $P_m[U^{-1}]$. Conclude that $\dim S = \sup\{d(m) - d(m-1) | 1 \leq m \leq \infty\}$. Thus if the $d(m) - d(m-1)$ are unbounded, then S has infinite dimension.

Show that S is Noetherian by proving the following lemma and checking its hypotheses.

Lemma 9.4. *Let S be a ring such that for every maximal ideal $P \subset S$ the local ring S_P is Noetherian. If for every element $s \in S$ there are only finitely many maximal ideals containing s, then S is Noetherian.*

10
The Principal Ideal Theorem and Systems of Parameters

In this chapter all rings will be assumed to be Noetherian.

It is elementary that a principal prime ideal in a Noetherian ring can have codimension at most 1. A sharper statement is this: Any prime properly contained in a proper principal ideal has codimension 0. *Proof*: If on the contrary, $Q \subsetneqq P \subsetneqq (x)$ in a ring R, with P and Q prime, then factoring out Q we can assume that $Q = 0$, and thus that R is a domain. If $y \in P$, then $y = ax$ for some a, and since $x \notin P$ it follows that $a \in P$; thus $P = xP$. By Corollary 4.7, $(1 - b)P = 0$ for some $b \in (x)$. Since R is a domain, we must have $b = 1$, so (x) is not proper, a contradiction.

The mainspring of the above argument is Nakayama's lemma, here in the guise of Corollary 4.7. A subtler application of Nakayama's lemma yields Krull's principal ideal theorem (PIT) [1928], a cornerstone of dimension theory for Noetherian rings. Krull's theorem extends the above remark from principal ideals to primes minimal over principal ideals. Geometrically, the result encapsulates and generalizes the dimension-theoretic side of the implicit function theorem in complex analysis, as we noted in Chapter 8. The principal ideal theorem says that even the most complex polynomial condition on the points of an algebraic variety is satisfied in codimension 1 if it can be satisfied at all.

Krull was the first to show that not only primary decomposition, but also a great deal of the geometric theory of the polynomial ring, could be carried over to the general Noetherian case. He deserves credit, after Emmy Noether, for making the theory of Noetherian rings viable. We give his beautiful proof of the principal ideal theorem; the result had

been proved by Kronecker in the 1880s for polynomial rings by a difficult elimination-theoretic argument, and a geometric version was known (if not proved) well before that, as part of the method of proof known in enumerative geometry as "counting constants."

Theorem 10.1 (First version of the Principal Ideal Theorem). *If $x \in R$, and P is minimal among primes of R containing x, then* $\operatorname{codim} P \le 1$.

Note that this result is vacuous if the element x is a unit, since then no primes contain it.

For the proof we shall freely use the equivalences in Corollary 2.19, which characterizes primes minimal over a given ideal. The proof uses an idea from primary decomposition. Recall from Chapter 3 that if $Q \subset R$ is a prime ideal, then the **nth symbolic power** $Q^{(n)} = \{r \in R | sr \in Q^n$ for some $s \in R, s \notin Q\}$ is the preimage in R of the nth power of the localized ideal Q_Q in R_Q. The elements outside Q are nonzerodivisors mod $Q^{(n)}$, and on localization we get $(Q^{(n)})_Q = (Q_Q)^n$.

Proof. We shall show that if Q is any prime ideal with $Q \subsetneq P$, then R_Q has dimension 0, so $\operatorname{codim} Q = 0$. This shows that $\operatorname{codim} P \le 1$.

Replacing R by R_P we may assume that P is maximal, and we have ideals as in Figure 10.1. Since P is minimal over (x), the ring $R/(x)$ is Artinian by Theorem 2.14. Thus the descending chain $(x) + Q^{(n)}$ stabilizes, say with $Q^{(n)} \subset (x) + Q^{(n+1)}$. It follows that for any $f \in Q^{(n)}$ we may write $f = ax + g$ with $g \in Q^{(n+1)}$. This implies that $ax \in Q^{(n)}$. Since P is minimal over (x), we have $x \notin Q$, so $a \in Q^{(n)}$.

From this we see that $Q^{(n)} = (x)Q^{(n)} + Q^{(n+1)}$. Since $x \in P$, Nakayama's lemma, Corollary 4.8a, implies $Q^{(n)} = Q^{(n+1)}$. A second application of Nakayama's lemma, this time in R_Q, yields $(Q_Q)^n = 0$, so R_Q has dimension 0 as claimed. □

This version of the principal ideal theorem can serve as the first step in an induction, yielding a result about primes minimal over an ideal with many generators.

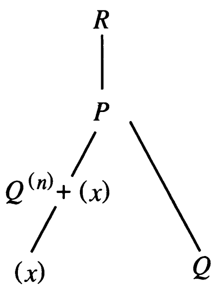


FIGURE 10.1.

Theorem 10.2 (Final version of the Principal Ideal Theorem). *If $x_1, \dots, x_c \in R$, and P is minimal among primes of R containing $x_1, \dots, x_c$, then* $\operatorname{codim} P \le c$.

Proof. We may suppose by localizing that P is the unique maximal ideal of R. By Corollary 2.17, P is nilpotent modulo the ideal $(x_1, \dots, x_c)$. Let P_1 be a prime such that $P \supset P_1$, with no prime between. We shall show that P_1 is minimal over an ideal generated by $c-1$ elements. By the inductive hypothesis, $\operatorname{codim} P_1 \le c-1$, and this suffices.

By hypothesis, P_1 cannot contain all the x_i; for definiteness, suppose $x_1 \notin P_1$. Thus P is minimal over (P_1, x_1), so P and in particular all the x_i are nilpotent mod (P_1, x_1). This means that for suitable n we can find elements $a_i \in R$ and $y_i \in P_1$ such that

$$x_i^n = a_i x_1 + y_i, \quad i = 2, \dots, c.$$

We claim that P_1 is minimal among primes containing $y_2, \dots, y_c$. Indeed, P is nilpotent mod $(x_1, y_2, \dots, y_c)$ so, by Theorem 10.1 the image of P in $R/(y_2, \dots, y_c)$ has codimension at most 1. Thus the image of P_1 in $R/(y_2, \dots, y_c)$ has codimension 0, the desired result. □

As a first consequence we have a strong descending chain condition on prime ideals in a Noetherian ring.

Corollary 10.3. *The prime ideals in a Noetherian ring satisfy the descending chain condition, with the length of a chain of primes descending from a prime P bounded by the number of generators of P.*

This allows us to take a major step toward Theorem A of Chapter 8.

Corollary 10.4. *The ideal $(x_1, \dots, x_c) \subset k[x_1, \dots, x_r]$ has codimension c.*

But note that it does not quite suffice to compute the dimension of the polynomial ring (we shall finish the job in Corollary 10.13).

There is a useful converse to the PIT, as follows.

Corollary 10.5 (Converse of the PIT). *Any prime P of codimension c is minimal over an ideal generated by c elements.*

Proof. Inductively, with $0 \le r < c$, having chosen $x_1, \dots, x_r \in P$ to generate an ideal of codimension r, it suffices to choose $x_{r+1} \in P$ but not in any of the finitely many primes minimal over $(x_1, \dots, x_r)$; this is possible by prime avoidance, Lemma 3.3, and Theorem 10.2. □

Another easy consequence of the principal ideal theorem is an improved characterization of factoriality.

Corollary 10.6. *A domain R is factorial if every codimension* 1 *prime of R is principal.*

Proof. By the principal ideal theorem and its converse, the codimension 1 primes are precisely the primes minimal over principal ideals, so we may apply Proposition 3.11b. □

The principal ideal theorem is a first assertion giving an interesting geometric conclusion from some hypothesis on the form of the generators of an ideal. There have been many extensions of this way to extract geometry from algebra. For example, there are similar bounds on the codimension of determinantal ideals (see Exercise 10.9).

A very general question in this direction is: Given an ideal I in a ring R, what is the maximal possible codimension for ideals of the form $\varphi(I)S$, where $\varphi : R \to S$ is a ring homomorphism such that $\varphi(I)S \neq S$? The supremum of such codimensions is called the "superheight" of I (see Hochster [1976]; to understand the name, recall that "height" is a synonym for "codimension"). The principal ideal theorem may be restated by saying that if $I = (x) \subset \mathbf{Z}[x] = R$, then superheight $I = 1$. It is known (Serre [1957]) that the superheight of I is equal to the codimension of I when I is prime and R is a regular local ring (the definition is given later in this chapter) or when R/I has a projective resolution over R whose length is codim I (see, for example, Hochster [1987]). The restriction to prime ideals must be made only to avoid rather trivial phenomena. However, there are many prime ideals in more complicated rings for which the superheight is strictly larger than the codimension; see Exercise 10.6, for one such instance, and Koh [1988] for further work on this idea.

10.1 Systems of Parameters and Parameter Ideals

Summarizing much of what we have done, we get another characterization of the dimension of a local ring:

Corollary 10.7. *If R is a local ring with maximal ideal $\mathfrak{m}$, then* $\dim R$ *is the smallest number d such that there exist d elements $x_1, \dots, x_d \in \mathfrak{m}$ with $\mathfrak{m}^n \subset (x_1, \dots, x_d)$ for $n \gg 0$.*

Proof. If $\mathfrak{m}^n \subset (x_1, \dots, x_d) \subset \mathfrak{m}$, then $\mathfrak{m}$ is minimal among primes over $(x_1, \dots, x_d)$ and $\dim R \leq d$ by the PIT.

On the other hand, we may find elements $x_1, \dots, x_d$ with $d = \dim R$ such that $\mathfrak{m}$ is a minimal prime containing $(x_1, \dots, x_d)$ by the converse of the PIT. But then $R/(x_1, \dots, x_d)$ has only one prime ideal, which must be nilpotent by Corollary 2.12, and we are done. □

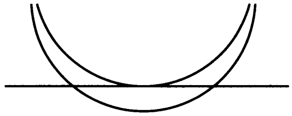


FIGURE 10.2. The functions y, $x^2 - y$ form a system of parameters for $k[x,y]_{(x,y)}$. Only finitely many points lie on the intersection of the level sets $x^2 - y = \delta$, $y = \varepsilon$ for small δ and ε.

Corollary 10.7 turns out to be so important that we codify the notions it uses: An ideal $\mathfrak{q} \subset \mathfrak{m}$ such that $R/\mathfrak{q}$ has finite length (equivalently, $\mathfrak{m}^n \subset \mathfrak{q}$ for $n \gg 0$) is called a **parameter ideal for *R***, and a sequence of elements $x_1, \ldots, x_d$ as in Corollary 10.7 is called a **system of parameters** for R.

Geometrically, if R is the local ring of a point p on an algebraic variety X, a system of parameters in R is a sort of local coordinate system for X around p, in the sense that the values of the functions x_i determine points near p up to a finite ambiguity, as in Figure 10.2. Systems of parameters can be characterized as the smallest sets of elements with this sufficiency property, or as those sets having a certain algebraic independence property (see Exercise 14.8).

More generally, if M is any finitely generated module over the local ring $(R, \mathfrak{m})$, then we say that an ideal $\mathfrak{q} \subset \mathfrak{m}$ is a parameter ideal for M if $M/\mathfrak{q}M$ has finite length. By Corollary 2.17, this is true iff a power of $\mathfrak{m}$ annihilates $M/\mathfrak{q}M$. Recalling that ann M denotes the annihilator of M, and using Nakayama's lemma, we may write this condition as $\operatorname{rad}(\operatorname{ann}(M/\mathfrak{q}M)) = \mathfrak{m}$.

The next result shows that parameter ideals for modules are connected with dimension theory in the general case just as in the special case above.

Proposition 10.8. *If M is a finitely generated R-module, and $\mathfrak{q}$ is any ideal of R then* $\operatorname{rad}(\operatorname{ann}(M/\mathfrak{q}M)) = \operatorname{rad}(\mathfrak{q} + \operatorname{ann} M)$. *In particular, if R is a local ring with maximal ideal $\mathfrak{m}$, then:*

a. $\mathfrak{q}$ is a parameter ideal for M iff $(\mathfrak{q} + \operatorname{ann} M) \supset \mathfrak{m}^n$ for $n \gg 0$ iff $\mathfrak{q}$ is a parameter ideal for $R/(\operatorname{ann} M)$.

b. Given a short exact sequence of modules

$$0 \to M' \to M \to M'' \to 0,$$

an ideal $\mathfrak{q}$ is a parameter ideal for M iff $\mathfrak{q}$ is a parameter ideal for M' and M''.

c. $\dim M$ *is the least number d such that there is a parameter ideal for M generated by d elements.*

Proof. To prove the equality on radicals it suffices, by Corollary 2.12, to show that a prime P of R contains $\operatorname{ann}(M/\mathfrak{q}M)$ iff P contains $\mathfrak{q} + \operatorname{ann} M$. By Proposition 2.1, $P \supset \operatorname{ann}(M/\mathfrak{q}M)$ iff $(M/\mathfrak{q}M)_P \neq 0$. By Nakayama's

lemma $(M/\mathfrak{q}M)_P = M_P/\mathfrak{q}_P M_P \neq 0$ iff $M_P \neq 0$ and $\mathfrak{q}_P \subset P_P$. By Proposition 2.1 these conditions are satisfied iff P contains both $\operatorname{ann} M$ and $\mathfrak{q}$; that is, iff $P \supset \mathfrak{q} + \operatorname{ann} M$.

a. Note that the annihilator of $R/(\operatorname{ann} M)/\mathfrak{q}(R/(\operatorname{ann} M))$ is $\mathfrak{q} + \operatorname{ann} M$. Also, since $\mathfrak{m}$ is the unique maximal ideal of R, we have $\operatorname{rad}(\mathfrak{q} + \operatorname{ann} M) = \mathfrak{m}$ iff $\mathfrak{q} + \operatorname{ann} M$ contains a power of $\mathfrak{m}$. The radical formula just established thus proves the equivalence of the three assertions.

b. If $\mathfrak{q}$ is a parameter ideal for M, then it is for M' and M'' because their annihilators contain the annihilator of M. The converse follows from the induced exact sequence

$$M'/\mathfrak{q}M' \to M/\mathfrak{q}M \to M''/\mathfrak{q}M'' \to 0$$

which shows that if $M'/\mathfrak{q}M'$ and $M''/\mathfrak{q}M''$ have finite length then $M/\mathfrak{q}M$ does also.

c. By definition $\dim M = \dim R/(\operatorname{ann} M)$, so conclusion c follows from a and Corollary 10.7. □

The principal ideal theorem deals with codimension rather than dimension, and this is occasionally a nuisance. However, a version with dimension follows in the local case. The local assumption is necessary for fairly trivial reasons; see Exercise 10.8.

Corollary 10.9. *If $(R, \mathfrak{m})$ is a local ring and M is a finitely generated R-module, then for any $x \in \mathfrak{m}$ we have*

$$\dim M/xM \geq \dim M - 1.$$

Proof. To say that $\dim M/xM = d$ means that $\dim R/\operatorname{ann}(M/xM)$ is a ring of dimension d. By Proposition 10.8a, there is a parameter ideal for M/xM generated by d elements $x_1, \ldots, x_d$. But this means that $M/(x, x_1, \ldots, x_d)M$ has finite length, so $(x, x_1, \ldots, x_d)$ is a parameter ideal for M, and $\dim M \leq 1 + d$ as required. □

10.2 Dimension of Base and Fiber

Corollary 10.7 also yields a "superheight" result: If R is local with maximal ideal $\mathfrak{m}$ and S is an R-algebra with $\mathfrak{m}S \neq S$, then $\operatorname{codim} \mathfrak{m}S \leq \operatorname{codim} \mathfrak{m}$. (*Proof*: If $x_1, \ldots, x_d$ is a system of parameters in R, then any prime minimal over $\mathfrak{m}S$ is minimal over $(x_1, \ldots, x_d)S$.) The inequality in the following theorem, which corresponds to part of Axiom D2 from Chapter 8, gives an extremely useful extension of this idea.

Theorem 10.10. *If* $(R, \mathfrak{m}) \to (S, \mathfrak{n})$ *is a map of local rings, then*

$$\dim S \leq \dim R + \dim S/\mathfrak{m}S,$$

with equality if S *is flat as an* R*-module.*

Equality in Theorem 10.10 is a much weaker condition than flatness, as the example in Figure 10.4 (which is not flat) shows. However, under strong hypotheses on R and S, flatness is equivalent to a statement about fiber dimensions; see Theorem 18.16. Some strengthenings of Theorem 10.10, in the case where S is a localization of a finitely generated R-algebra, that let one compute the dimension of S without a flatness hypothesis, are given in Corollary 13.5 and Theorem 13.8.

Proof. Write $d = \dim R$, and $e = \dim S/\mathfrak{m}S$. By Corollary 10.7 there exist $x_1, \ldots, x_d \in \mathfrak{m}$ such that $\mathfrak{m}^s \subset (x_1, \ldots, x_d)$ for $s \gg 0$ and $y_1, \ldots, y_e \in S$ such that $\mathfrak{n}^t \subset \mathfrak{m}S + (y_1, \ldots, y_e)$ for $t \gg 0$. Thus

$$\begin{aligned} \mathfrak{n}^{st} &\subset (\mathfrak{m}S + (y_1, \ldots, y_e))^s \\ &\subset \mathfrak{m}^s S + (y_1, \ldots, y_e) \\ &\subset (x_1, \ldots, x_d, y_1, \ldots, y_e)S, \end{aligned}$$

and by the principal ideal theorem $\dim S \leq d + e$.

Now suppose that S is flat as an R-module. We must show that $\dim S \geq \dim R + \dim S/\mathfrak{m}S$. Let Q be a prime of S, minimal over $\mathfrak{m}S$, such that $\dim Q = \dim S/\mathfrak{m}S$. We have

$$\dim S \geq \dim Q + \operatorname{codim} Q = \dim S/\mathfrak{m}S + \operatorname{codim} Q,$$

so it suffices to show that codim $Q \geq \dim R$. Writing φ for the given map $R \to S$, we have $\varphi^{-1}Q = \mathfrak{m}$. Thus it suffices to show that given a chain of primes $\mathfrak{m} \supset P_1 \supset \cdots$ of R, descending from $\mathfrak{m}$, there is a chain of primes $Q \supset Q_1 \supset \cdots$ of S with $\varphi^{-1}Q_i = P_i$. This is the content of the following "going down" lemma. □

Lemma 10.11 (Going Down for flat extensions). *Suppose that* $\varphi : R \to S$ *is a map of rings such that* S *is flat as an* R*-module. If* $P \supset P'$ *are primes of* R *and* Q *is a prime of* S *with* $\varphi^{-1}Q = P$*, then there exists a prime* Q' *of* S *contained in* Q *such that* $\varphi^{-1}Q' = P'$ *as in Figure* 10.3. *In fact* Q' *may be taken to be any prime of* S *contained in* Q *and minimal over* $P'S$.

The proof uses a fundamental lemma from the theory of primary decomposition. An alternative (giving a slightly weaker result) is described in Exercise 10.7.

Proof. Since $P'S \subset Q$, we may find a prime Q' contained in Q and minimal over $P'S$ (here we use the fact that the intersection of a descending

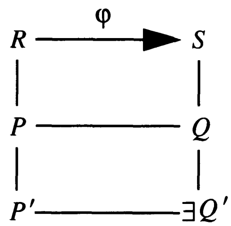


FIGURE 10.3.

chain of primes is prime). Since tensoring preserves flatness (Exercise 6.6a), $S/P'S = S \otimes R/P'$ is flat over R/P'. Replacing R by R/P' and S by $S/P'S$, we may reduce to the case $P' = 0$. Since S is flat over R, every nonzerodivisor in R—that is, every nonzero element of R—is a nonzerodivisor on S (Corollary 6.3). Since Q' is a minimal prime of S, it consists of zerodivisors on S by Theorem 3.1, proving that $\varphi^{-1}(Q') = 0$ as required. □

For another case in which "going down" holds, see Theorem 13.9.

It is not hard to translate "going down" into an interesting geometric statement: Suppose that $\psi : X \to Y$ is a map of affine varieties over an algebraically closed field with induced map $\varphi : R = A(Y) \to S = A(X)$. Suppose the prime Q corresponds to a subvariety Z of X, and that W is a subvariety of Y with $\psi(Z) \subset W$. If "going down" holds between R and S, there must be a subvariety $V \supset Z$ in X whose image under ψ is dense in W. An example where this fails is given in Figure 10.4.

A slightly more careful analysis leads to a geometric result extending this one: If $\psi : X \to Y$ is a map of affine varieties such that the induced map $\varphi : R := A(Y) \to S := A(X)$ makes S a flat R-module, then the map ψ is **open** in the sense that it carries open sets to open sets. See, for example, Hartshorne [1977, III, Ex. 9.1]. This is one of several ways in which flat morphisms behave like submersive maps of manifolds.

The next two corollaries are applications of Theorem 10.10.

Corollary 10.12. *If $R, \mathfrak{m}$ is a local ring and $\hat{R}$ its completion at $\mathfrak{m}$, then*

$$\dim \hat{R} = \dim R.$$

Proof. $\hat{R}$ is flat over R by Theorem 7.2, and the fiber $\hat{R}/\mathfrak{m}\hat{R}$ is the residue class field, which has dimension 0. □

Corollary 10.13.

a. If k is a field, then

$$\dim k[x_1, \ldots, x_r] = r.$$

b. More generally, if $R[x]$ is the polynomial ring in one variable over R, then

$$\dim R[x] = 1 + \dim R.$$

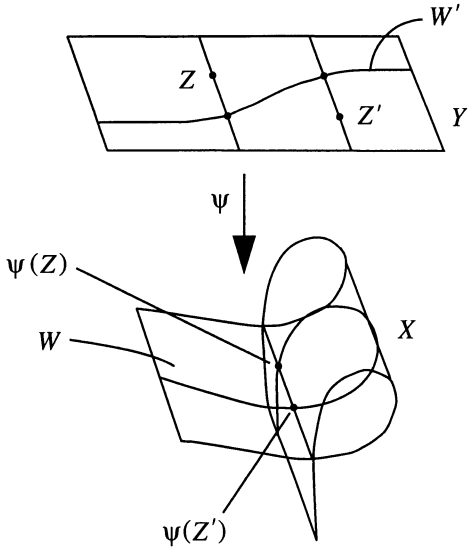


FIGURE 10.4. $\psi^{-1}(W)$ has components W', Z, and Z'. No irreducible variety containing Z maps to a dense subset of W.

c. *Furthermore, if P is a prime ideal of R then there are prime ideals Q of $R[x]$ contracting to P, and for a maximal such ideal we have*

$$\dim R[x]_Q = 1 + \dim R_P.$$

Proof. Part a follows from part b by induction on r.

For part b, first note that given a chain of primes $P_1 \subset \cdots \subset P_d$ of R we get a longer chain of primes $P_1R[x] \subset \cdots \subset P_dR[x] \subset P_dR[x] + (x)$ in $R[x]$, so $\dim R[x] \geq 1 + \dim R$. The other inequality follows from part c, since a maximal ideal Q of $R[x]$ is certainly maximal among primes meeting R in $P = Q \cap R$. Thus, it suffices to prove part c.

First we check the result in the case where R is a field and $P = 0$. In this case any prime ideal of $R[x]$ contracts to 0 in R, so we take Q to be any maximal ideal of $R[x]$. Since 0 is not a maximal ideal of $R[x]$, we have $Q \neq 0$, and $\operatorname{codim} Q \geq 1$. On the other hand, $R[x]$ is a principal ideal domain, so Q is principal and $\operatorname{codim} Q \leq 1$ by the principal ideal theorem (or by the elementary argument given at the beginning of this chapter). Thus, $\dim R[x]_Q = 1 = 1 + \dim R_P$, as required.

In the general case, $PR[x]$ is a prime ideal of $R[x]$ and $PR[x] \cap R = P$, proving the first statement of c. We may replace R by R_P and assume that R is local with maximal ideal P. Let Q be a maximal ideal of $R[x]$ containing P; it follows that $Q \cap R = P$, and we must show that codim $Q = 1 + \operatorname{codim} P$.

If $P_0 \subset \cdots \subset P_d = P$ is a chain of primes in R, then $P_0R[x] \subset \cdots \subset P_dR[x]$ is a chain of primes in $R[x]$ of the same length. By the previous

case, the maximal ideal $Q/PR[x]$ in

$$R[x]/PR[x] = (R/P)[x]$$

has codimension 1, so $\operatorname{codim} Q \geq d+1$.

Let $k = R_P/PR_P$ be the residue class field of R at P. We may apply Theorem 10.10 to get

$$\begin{aligned}\dim R[x]_Q &\leq \dim R_P + \dim R[x]_Q/PR[x]_Q \\ &\leq \dim R_P + \dim k[x] \\ &\leq \dim R_P + 1,\end{aligned}$$

completing the proof. □

For more on primes in polynomial rings, see Exercises 10.2 and 13.6.

10.3 Regular Local Rings

The principal ideal theorem gives us an interesting inequality connecting the codimension of an ideal with the number of its generators. It is often interesting to ask in such cases, "What happens in the case of equality?" Restricting the question to the case of the maximal ideal of a local ring, this strategy leads to a big payoff in the theory of regular local rings.

Suppose that $(R, \mathfrak{m})$ is a local ring of dimension d. The principal ideal theorem shows that $\mathfrak{m}$ cannot be generated by fewer than d elements. Following Krull, R is called **regular** if $\mathfrak{m}$ can be generated by exactly d elements. By Nakayama's lemma a collection of elements generates $\mathfrak{m}$ iff the images of these elements generate the $(R/\mathfrak{m})$-vector space $\mathfrak{m}/\mathfrak{m}^2$, so this vector space has dimension d iff R is regular, and then every minimal system of generators of $\mathfrak{m}$ has d elements. Such a minimal system of generators is a system of parameters for R; it is called a **regular system of parameters**.

Examples of regular local rings are $k[x_1, \ldots, x_d]_{(x_1,\ldots,x_d)}$ and the power series ring $k[[x_1, \ldots, x_d]]$, where k is a field. In both of these cases the sequence of variables forms a regular system of parameters. If p is a prime integer, then $\mathbf{Z}_{(p)}[x_1, \ldots, x_n]_{(p,x_1,\ldots,x_n)}$, and $\hat{\mathbf{Z}}_{(p)}[[x_1, \ldots, x_n]]$ are also regular local rings. (What is a regular system of parameters?)

Regular local rings occupy center stage in algebraic geometry, since, as Zariski realized, they correspond to nonsingular points on algebraic varieties. The simplest case is this: Over the complex numbers, an algebraic variety $X \subset \mathbf{C}^m$ looks like a complex analytic submanifold of $\mathbf{C}^m$ locally near a point x iff the localization of $A(X)$ at the prime ideal corresponding to x is a regular local ring. We shall prove an equivalent result in Theorem 16.19. (History here is not what one might expect. Krull made the definition [1937] a few years before Zariski proved the theorem [1940, 1947].

Of course some suggestive geometric cases were familiar to Krull.) The following is the tip of an iceberg of results showing that regular local rings are well behaved.

Corollary 10.14. *If R is a regular local ring, then R is an integral domain.*

Proof. Let R be a regular local ring with maximal ideal $\mathfrak{m}$. We do induction on $\dim R$. In case $\dim R = 0$, we must have $\mathfrak{m} = 0$, so R is a field, and the result is trivial. Thus we may suppose $\dim R = d > 0$.

By Nakayama's lemma, we have $\mathfrak{m}^2 \neq \mathfrak{m}$, so by prime avoidance (Lemma 3.3) and the finiteness of the set of minimal primes of R, we may find an element $x \in \mathfrak{m}$ that is outside the minimal primes of R, and also outside $\mathfrak{m}^2$. Set $S = R/(x)$, and let $\mathfrak{n} = \mathfrak{m}S$ be the maximal ideal of S. By the choice of x we have $\dim S < \dim R$, so $\dim S = d - 1$ by Corollary 10.9. Also, $\mathfrak{n}/\mathfrak{n}^2 = \mathfrak{m}/(\mathfrak{m}^2 + (x))$ is a proper homomorphic image of $\mathfrak{m}/\mathfrak{m}^2$, so it can be generated by $(d-1)$ elements. By Nakayama's lemma, $\mathfrak{n}$ can be generated by $(d-1)$ elements, so S is regular of dimension $d-1$. By induction S is a domain; that is, (x) is a prime ideal. Since we chose x outside the minimal primes, (x) is not a minimal prime of R. Thus (x) contains some minimal prime ideal Q of R.

If $y \in Q$ is any element, then we may write $y = ax$ for some $a \in R$. Since x is not in Q, we must have $a \in Q$. This shows that $Q = xQ$. It follows that $\mathfrak{m}Q = Q$, so by Nakayama's lemma $Q = 0$, and R is a domain as required. □

In general it is extremely difficult to prove that a given ideal of polynomials is prime; the simple idea of the proof just given is the basis of one of the most powerful methods known for doing so, Hochster's method of "principal radical systems" (Hochster [1976]). Another method, using Serre's criterion for normality, is described in Chapter 18 (Theorem 18.15).

Corollary 10.14 has an extension that we shall use many times. It is best stated in terms of a definition that generalizes the notion of a nonzerodivisor:

A sequence of elements $x_1, \ldots, x_d$ in a ring R is called an ***R*-sequence** (or **regular sequence** on R) if the ideal $(x_1, \ldots, x_d)$ is proper and for each i, the image of x_{i+1} is a nonzerodivisor in $R/(x_1, \ldots, x_i)$.

Corollary 10.15. *If $x_1, \ldots, x_d$ is a regular system of parameters in a regular local ring R, then $x_1, \ldots, x_d$ is an R-sequence.*

Proof. For each i the ring $R/(x_1, \ldots, x_i)$ is a regular local ring, and thus an integral domain by Corollary 10.14. The image of x_{i+1} in this domain must be nonzero, since the maximal ideal of R could otherwise be generated by fewer elements. □

In our earlier examples of regular local rings containing a field, the regular systems of parameters were systems of indeterminates. This is not true in

general; for example, if k is any field, then $(k[x,y]/(y^2-x(x-1)(x+1)))_{(x,y)}$ is a regular local ring—the dimension is 1, and the maximal ideal is generated by x. But if the characteristic of k is not 2, this ring is very different than $k[x]_{(x)}$; for example, its quotient field is not isomorphic to $k(x)$. However, just as all smooth manifolds of a given dimension look alike in a sufficiently small neighborhood, complete local rings do not vary very much.

Proposition 10.16. *Suppose R is a complete regular local ring of dimension d with residue class field k. If R contains a field, then $R \cong k[[x_1,\ldots,x_d]]$, and the isomorphism can be chosen to send the variables x_i to any given regular system of parameters in R.*

Proof. By the Cohen structure theorem, R contains a copy of its residue field k. If $y_1,\ldots,y_d$ is a regular system of parameters in R, then by Theorem 7.16 there is a surjective ring homomorphism $\varphi : k[[x_1,\ldots,x_d]] \to R$, sending x_i to y_i. Since $k[[x_1,\ldots,x_d]]$ is an integral domain of dimension d, any proper homomorphic image of $k[[x_1,\ldots,x_d]]$ has dimension less than d. Thus, φ is an isomorphism. □

Beyond this it is known, for example, that the elements of a regular local ring have unique factorizations into primes, that localizations of regular local rings are again regular, and much more $\cdots$. We shall prove some of these things in Chapter 19, after we have homological tools at our disposal.

10.4 Exercises

Exercise 10.1: Let R be a Noetherian ring, and let x be an indeterminate. Show that $\dim R[x,x^{-1}] = 1 + \dim R$.

Exercise 10.2 (Prime ideals in a polynomial ring):* Let R be a Noetherian ring. This exercise refines the formula $\dim R[x] = 1 + \dim R$. Suppose that P is a prime ideal of R of codimension c. Show that the prime ideals $Q \subset R[x]$ that intersect R in P are all of the following two kinds, with codimension as shown:

a. $Q = PR[x]$. In this case $\operatorname{codim} Q = c$.

b. $Q \supsetneqq PR[x]$. In this case $\operatorname{codim} Q = c+1$, and there is a polynomial $f(x) \in Q$ with leading coefficient not in P such that
$$Q = \{g \in R[x] \mid \text{for some } a \in R-P,\ ag \in PR[x] + (f)\}.$$

For each prime $P \subset R$ there are infinitely many primes in $R[x]$ of type b. See also Exercise 13.6.

Exercise 10.3: Let k be a field. Show that the ring $k[x] \times k[x]$ contains a principal prime ideal of codimension 1, although it is not a domain. (By the argument of Corollary 10.14, there is no such example in a local ring.)

Exercise 10.4:* Let a, b be a regular sequence in a domain R, and let $S = R[x]$ be the polynomial ring in one variable over R. Show that $ax - b$ is a prime of S. (See Exercise 17.2 for a sort of converse. Geometers will recognize this as a very weak version of a theorem of Bertini.)

Exercise 10.5: If $R = k[t]_{(t)}$ and S is a domain containing R, then S is torsion-free as an R-module and thus, as we have seen, flat, so the inequality of Theorem 10.10 is an equality. Show, however, that we may have strict inequality already in the case where $R = k[s,t]_{(s,t)} \subset S = k[s,t/s]_{(s,t/s)}$, as in Exercise 6.9. Note that there is an open dense set of points p of the blowup in Exercise 6.9 such that the equality holds in Theorem 10.10 if we take S to be the localization of $k[s,t/s]$ at the maximal ideal corresponding to p, and R to be the localization of $k[s,t]$ at the preimage of this maximal ideal. This phenomenon is general; see Corollary 14.5.

Exercise 10.6: We mentioned that if P is a prime ideal in a regular local ring R and if $R \to S$ is a map of local rings, then $\operatorname{codim} PS \leq \operatorname{codim} P$. Here is an example showing that this may fail when R is not regular: Let $R = k[x,y,s,t]/(xs - yt)$, and let $S = R/(x,y) = k[s,t]$. Let $P = (s,t) \subset R$. Prove that $\operatorname{codim} P = 1$, but $\operatorname{codim} PS = 2$, as shown in Figure 10.5.

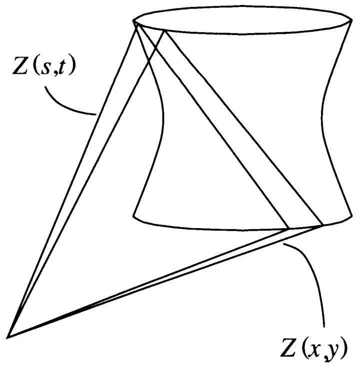


FIGURE 10.5. Two planes in 4-space meeting in a point lie on the cone over a quadric surface.

Exercise 10.7: Here is a weak "going down" statement without the hypothesis of flatness: Show that if R is an integral domain contained in the local ring (S, Q), then there is a minimal prime of S contracting to 0 in R. Give another proof of Lemma 10.11 using this in place of Theorem 3.1.

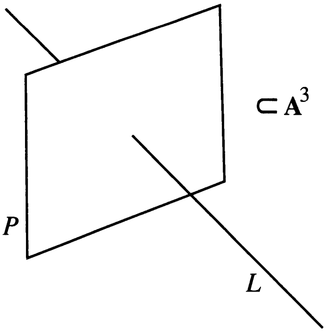


FIGURE 10.6. $X = P \cup L$, R is the coordinate ring of X.

Exercise 10.8: The condition that R be local is needed in Corollary 10.9 to avoid a rather trivial sort of counterexample. Find the one illustrated by Figure 10.6, where X has dimension 2, but a hyperplane section of X by a hyperplane parallel to P would only have dimension 0.

Determinantal Ideals

Exercise 10.9 Determinantal ideals and a generalized principal ideal theorem:* Let M be a $p \times q$ matrix with entries in a ring R (Noetherian as always). Recall that a $\boldsymbol{k \times k}$ **minor** of M is by definition the determinant of some $k \times k$ submatrix of M. Let $I_k(M)$ be the ideal generated by the $k \times k$ minors of M. If $k = 1$ and $p = 1$, the PIT bounds the codimension of a prime minimal over $I_k(M)$ by q. Macaulay [1916] generalized this to the case $k = p$ arbitrary, giving the bound $q - k + 1$. (It is amusing to note that Macaulay proved this generalized principal ideal theorem in the context of polynomial rings before Krull proved Theorem 10.2.) The following generalization to the case where p and k are both arbitrary was given around 1960 by Jack Eagon in his thesis. The proof is taken from a subsequent paper by Eagon and Northcott [1962].

Prove that if P is a prime minimal over $I_k(M)$, then P has codimension $\leq (p - k + 1)(q - k + 1)$ as follows:

a.* (Reduction) Show that if one of the entries of M is a unit, then $I_k(M) = I_{k-1}(M')$ where M' is a suitable $(p-1) \times (q-1)$ matrix.

b. Localize at P. Use induction on k to reduce to the case where all entries are in P and $k > 1$.

c. Now pass to $R[t]$, where t is a new indeterminate. If the $1,1$ entry of M was $x_{1,1}$, replace M by the new matrix M' whose $1,1$ entry is $x_{1,1}+t$ and whose other entries are the same as those of M. Show that $I_k(M') \subset P' := PR[t]$. Since P' has the same codimension as P, and M' has the $1,1$ entry not contained in P', the problem reduces to showing that P' is minimal over $I_k(M')$.

d. Prove that $P' + (t)$ is minimal over $I_k(M') + (t)$.

e. Suppose that $P' \supset Q \supset I_k(M')$, with Q a prime of $R[t]$. Show that $P' + (t)$ is minimal over $Q + (t)$. Deduce that $P' = Q$ by the principal ideal theorem in $R[t]/Q$.

Exercise 10.10:* Let R be any Noetherian ring, and let $M = (x_{ij})$ be a $p \times q$ matrix of indeterminates over the polynomial ring $R[x_{ij}]$. Show that $I_k(M)$ has codimension $(p - k + 1)(q - k + 1)$.

Hilbert Series of a Graded Module

Let k be a field, let S be a polynomial ring over k, and let M be a finitely generated graded S-module. In the Introduction we saw that if the variables of S all have degree 1, then the Hilbert function $H_M(n) = \dim_k M_n$ agrees with a polynomial function of n for large n. This is not true when the variables have different degrees, as the following exercises show. In this case $H_M(n)$ still agrees with a "periodic polynomial," but it is often more convenient to use the Hilbert series instead.

Exercise 10.11: Let S be the graded polynomial ring $k[x_1, x_2]$, where we give x_i degree i.

a. Show that $H_S(n) = \lfloor n/2 \rfloor + 1$. Show that this does not agree with a polynomial function in n, even for $n \gg 0$.

b. Show that $H_S(2n)$ and $H_S(2n + 1)$ are both polynomial functions of n for $n \geq 0$. Show also that the **Hilbert series** $\sum_{n \geq 0} H_A(n)t^n$ is a rational function in t with denominator $(1 - t)(1 - t^2)$.

Exercise 10.12: Let $S = k[x_1, \ldots, x_r]$, where x_i is an indeterminate of degree d_i. Set $q(t) = \Pi_{i=1}^r (1 - t^{d_i})$. The Hilbert series of M is defined to be the formal power series in one variable t given by $h_M(t) := \sum_{n \geq 0} H_M(n)t^n$.

a. Show that $h_M(t)$ is a rational function of t, and that in fact $h_M(t)$ may be written as a polynomial divided by $\Pi_{i=1}^u (1 - t^{d_i})$; that is, it is a rational function with poles only at roots of unity.

b. Show that there is a number d (which may be taken to be the least common multiple of the degrees of the d_i) such that for each s, $H_M(dn + s)$ agrees with a polynomial in n for all $n \gg 0$; that is, $H_M(n)$ is a "polynomial with periodic coefficients."

c. Imitating the proof that the Hilbert function agrees with a polynomial for large n in the classical case, show that the Hilbert series of M is a rational function.

Exercise 10.13: Continuing with the notation in Exercise 10.12:

a. Now suppose that M has a system of parameters $y_1, \ldots, y_m$, where y_i is a homogeneous element of degree $e_i > 0$. Suppose further that $y_1, \ldots, y_m$ is a **regular sequence** on M (that is, y_i is a nonzerodivisor on $M/(y_1, \ldots, y_{i-1})M$ for $i = 1, \ldots, m$; because of the grading the condition $(y_1, \ldots, y_m)M \neq M$ is automatic). Show that $h_M(t) = u(t)/s(t)$, where $s(t) = \Pi_{i=1}^m (1 - t^{e_i})$ and $u(t) = \sum u_i t^i$, where $u_i = H_{[M/(y_1,\ldots,y_m)M]}(i)$ gives the Hilbert function of the module $M/(y_1, \ldots, y_m)M$. Note that this is a module of finite length, so that $u(t)$ really is a polynomial, and that the coefficients of $u(t)$ are nonnegative integers.

b. Let $S = k[x_1, x_2]$, where both variables have degree 1. Compute the Hilbert series of the S-module $S/(x_1^2, x_1 x_2)$. Since x_2 is a system of parameters, the Hilbert series of S can be written with denominator $(1 - t)$; note that the numerator does not have positive coefficients.

c. Since the Hilbert series of M is a rational function defined over the integers by Exercise 10.12c, it makes sense to speak of the order of its pole at any number. Show that the dimension of M is the order of the pole of $h_M(t)$ at $t = 1$. (Hint: Use a system of parameters for M).

11
Dimension and Codimension One

A large part of classical algebraic geometry has to do with geometry "in codimension one": results about points in curves, curves in surfaces, and so forth. The commutative algebra of codimension one is correspondingly rich. In this chapter we digress from the presentation of dimension theory and use the results of Chapters 9 and 10 to analyze some codimension-1 phenomena. In particular, we shall study "invertible" modules; give a criterion for and some consequences of normality, including a bit of the theory of Dedekind domains; study the length of a one-dimensional ring modulo a principal ideal; and prove that the integral closure of a one-dimensional Noetherian domain is Noetherian.

Since normalization is such a well-behaved operation, a major strategy in commutative algebra is to analyze any ring by comparing it with its normalization. Our story begins with a description of one-dimensional normal local rings.

11.1 Discrete Valuation Rings

Recall that a one-dimensional local ring $(R, \mathfrak{m})$ is said to be regular iff its maximal ideal can be generated by one element. A generator for $\mathfrak{m}$ is called a **regular** (or **uniformizing**) **parameter** for R. Such rings have a very simple structure:

Proposition 11.1. *Let* $(R, \mathfrak{m})$ *be a regular local ring of dimension* 1. *If* π *is a regular parameter for* R, *then every element* t *of the quotient field*

$K(R)$ can be written uniquely in the form $t = u\pi^n$ with $n \in \mathbf{Z}$ and u a unit of R. In particular, every ideal of R is of the form (π^n), and R is a principal ideal domain.

It follows from this result that R is factorial; we shall prove that all regular local rings are factorial in Chapter 19.

Proof. Since R is a regular local ring, it is a domain by Corollary 10.14. If $s \in R$, we may by the Krull intersection theorem choose a representation $s = v\pi^m$ with $m \in \mathbf{Z}$ as large as possible such that $v \in R$. Since $v \notin (\pi) = \mathfrak{m}$, v must be a unit. If $t \in K(R)$, then writing $t = s_1/s_2$ with $s_i \in R$, and applying the result for s to s_1 and s_2, we see that t may be represented as $t = u\pi^n$ with $n \in \mathbf{Z}$ and $u \in R$ a unit. Such a representation is unique because if $u\pi^n = u'\pi^{n'}$, then $u/u' = \pi^{n'-n}$ is a unit of R, so $n' = n$ and $u = u'$.

The last two statements follow easily. □

Let R be a one-dimensional regular local ring. We may define a group homomorphism $\nu : K(R)^* \to \mathbf{Z}$, from the multiplicative group $K(R)^*$ of nonzero elements of the quotient field of R to $\mathbf{Z}$, as follows. If $t \in K(R)^*$, then by Proposition 11.1 there is a unique $n \in \mathbf{Z}$ such that $t = u\pi^n$ with u a unit of R. We set $\nu(t) = n$. The map ν is a **discrete valuation**. Note that if $s = v\pi^m$, then $s + t = (u\pi^{m-d} + v\pi^{n-d})\pi^d$ with $d = \min(m, n)$, so $\nu(s + t) \geq \min(\nu(s), \nu(t))$. In general, a **valuation** on a domain R is a group homomorphism ν from $K(R)^*$ to a totally ordered group G such that $\nu(r) \geq 0$ for $r \in R$, satisfying the inequality $\nu(a+b) \geq \min(\nu(a), \nu(b))$. The word *discrete* refers to the fact that $G = \mathbf{Z}$, the only discrete subgroup of the totally ordered group of real numbers under addition. The **valuation ring** of ν is the ring $S = \nu^{-1}\{g \in G | g \geq 0\}$. One-dimensional regular rings are usually called **discrete valuation rings**, or **DVRs**. For more on valuations, see Exercises 11.1–11.5.

Familiar examples of DVRs include the localization of a polynomial ring in one variable over a field $k[t]_{(t)}$ and its completion $k[[t]]$; and the arithmetic analogues of these rings, $\mathbf{Z}_{(p)}$ for any prime p and its completion $\hat{\mathbf{Z}}_p$, the ring of p-adic integers.

By Proposition 10.16, two complete DVRs that contain fields are isomorphic iff they have isomorphic residue class fields (and then they are isomorphic to $k[[x]]$, where k is the residue class field). On the other hand, DVRs that are not complete can be very different from one another, even when they are localizations of one-dimensional affine rings over a given algebraically closed field. To give a quick example we appeal to some facts from algebraic geometry. Suppose that $R = A_P$ and $R' = A'_{P'}$ are DVRs obtained by localizing the affine rings A and A' of nonsingular curves C and C' over an algebraically closed field k at primes P and P' corresponding to points $p \in C$ and $p' \in C'$, respectively. A basic argument from algebraic

geometry (Hartshorne [1977, Chapter I, section 6]) shows that C and C' can be embedded as open subsets of unique nonsingular projective curves $\bar{C}$ and $\bar{C}'$ and that if $R \cong R'$ as k-algebras, then $\bar{C} \cong \bar{C}'$ by an isomorphism carrying p to p'. If the genus of $\bar{C}$ is at least 2, then $\bar{C}$ has only finitely many automorphisms, so given p there are only finitely many points p' whose local rings are isomorphic to that of p. In particular, R is isomorphic to A_Q for only finitely many maximal ideals Q of A other than P. By Theorem 4.19 A is a Jacobson ring, and thus has infinitely many maximal ideals. So it has infinitely many isomorphism classes of localizations at maximal ideals.

11.2 Normal Rings and Serre's Criterion

In general, normalization has the effect of smoothing out certain irregularities in a variety; it is a step toward a "resolution of singularities." We shall show that a normal one-dimensional local Noetherian ring is a DVR; in particular, a normal one-dimensional variety is already nonsingular. Normalization gives a cheap and canonical process for resolution of singularities in dimension one. Since normalization commutes with localization, it follows from this that the localizations of a normal ring at codimension-1 primes are regular—that is, they are discrete valuation rings. This is called **regularity in codimension 1**. In the geometric setting, this means that a normal variety is nonsingular in codimension 1—that is, the singular locus is of codimension ≥ 2. Our first main result, Serre's criterion, explains the condition that must be added to make the converse of this statement true as well.

As an introduction, consider the case of a factorial domain. We have seen in Proposition 4.10 that any factorial domain R is normal. The heart of this argument is the following: Suppose r/s is integral over R and we wish to prove that $r/s \in R$, that is, that $r \in (s)$. If $(s) = R$ we are done, so we assume that (s) is a proper ideal. The argument given in Proposition 4.10 deduces from the integrality of r/s a relation of the form $ar = bs$. This shows that if $r \notin (s)$, then at least r is a zerodivisor modulo s and is contained in an associated prime of s. In the factorial case, this associated prime is generated by one of the prime elements dividing s, so we can divide both r and s by its generator and complete the proof by Noetherian induction (or, as in the version of Proposition 4.10, we could assume from the outset that r and s have no common factor, and derive a contradiction). Thus the condition we used could be stated as follows: *Associated primes of principal ideals are principal.* It is not hard to check that this condition is equivalent to factoriality (see Corollary 10.6). But a slightly weaker condition of this sort is equivalent to normality.

Theorem 11.2. *A Noetherian domain R is normal iff* $(*)$ *for every prime P of R associated to a principal ideal, P_P is principal.*

This terse statement deserves some amplification. First, the condition that P_P is principal implies (by the principal ideal theorem) that codim $P = 1$, so $(*)$ implies that (i) every associated prime to a principal ideal has codimension 1.

Further, if codim $P = 1$, so that R_P is one-dimensional, then P_P is principal iff R_P is a discrete valuation ring. Thus $(*)$ is equivalent to condition i along with condition ii: Every localization of R at a codimension-1 prime is a discrete valuation ring (regularity in codimension 1). The theorem asserts that normality is equivalent to conditions i and ii.

The one-dimensional local case of Theorem 11.2 says that a one-dimensional local domain is normal iff it is a discrete valuation ring (that is, iff its maximal ideal is principal).

Proof. We first show that the given conditions imply that R is normal. Since an intersection of normal domains with a common quotient field is obviously normal, it will be enough to show that R is the intersection of its localizations at primes associated to principal ideals. This is done by the following proposition. Since we shall soon want to apply it in a case where R is not a domain but only reduced, we prove it more generally for reduced rings. Here we need some terminology: We shall say that a prime of R is **associated to a nonzerodivisor** if it is an associated prime of a principal ideal generated by a nonzerodivisor.

Proposition 11.3. *If R is a reduced Noetherian ring, then an element $x \in K(R)$ belongs to R iff the image of x in $K(R)_P$ belongs to R_P for every prime P associated to a nonzerodivisor in R.*

Proof. Suppose that $a/u \in K(R)$, with $a, u \in R$ and u a nonzerodivisor. If $a/u \notin R$, then $a \notin (u)$. By Corollary 3.5, there is an associated prime P of (u) such that $a \notin (u)_P \subset R_P$. Thus $a/u \notin R_P$. □

Continuing the proof of Theorem 11.2, we next suppose that R is a normal domain and that P is a prime of R associated to a principal ideal (a); say P is the annihilator of $b \operatorname{mod}(a)$, with $b \in R - (a)$. We shall show that P_P is a principal ideal of R_P. Localizing if necessary, we may assume from the outset that R is local with maximal ideal P. Let K be the quotient field of R, and consider the set $P^{-1} := \{r \in K | rP \subset R\}$. We clearly have $P \subset P^{-1}P \subset R$, and since P is maximal, this leaves only the possibilities $P^{-1}P = P$ and $P^{-1}P = R$.

If $P^{-1}P = P$, then by Corollary 4.6 the elements of P^{-1} are integral over R. Since R is normal, $P^{-1} = R$. But $Pb \subset (a)$, so $b/a \in P^{-1} = R$, whence $b \in (a)$—contradicting our assumption.

Thus $P^{-1}P = R$; that is, P is invertible. Since R is local, we get $P \cong R$ by Theorem 11.6a, so P is principal. (The necessary special case of Theorem 11.6 is easy to do directly: Since R is local, $P^{-1}P = R$ implies that for some $r \in P^{-1}$ we have $rP = R$. Consequently, $P = Rr^{-1}$ is principal.) □

Corollary 11.4. *If R is a normal Noetherian domain, then R is the intersection of its localizations at codimension-1 primes.*

Proof. By Proposition 11.3 any ring is the intersection of its localizations at the primes associated to nonzerodivisors. If R is normal and P is a prime associated to a nonzerodivisor, then we have shown that P_P is principal. Thus P_P, and with it P, has codimension 1. □

The geometric version of Corollary 11.4 is quite useful. It says that if X is a normal variety and $Y \subset X$ is a subvariety of codimension at least 2, then any rational function on X regular on $X - Y$ extends to a regular function everywhere on X. Another version (that we shall not prove) of this is the **removable singularities theorem** of several complex variables: If X is a normal analytic variety of dimension at least 2, and $x \in X$ is a point (or more generally a codimension-2 subset), then any meromorphic function on X that is holomorphic outside x is holomorphic everywhere on X.

There is an important extension of the criterion of Theorem 11.2. First, a definition: A ring R is **normal** if it is reduced, and integrally closed in its total quotient ring. The **normalization** of a reduced ring R is the integral closure of R in its total quotient ring.

Serre noticed that conditions i and ii following Theorem 11.2 really have nothing to do with R being a domain, and that with small modifications they distinguish normal rings among all Noetherian rings. It turns out that a normal ring is a direct product of normal domains. A local or graded ring cannot be a nontrivial direct product, so in many "practical" cases, Theorem 11.5 gives a criterion that serves to distinguish normal domains.

It happens that it is generally rather hard to prove that a particular ring is a domain, that is, that an ideal is prime, while it is often not so hard to check the conditions of the criterion. For instance, condition i follows from the Cohen-Macaulay condition that we shall investigate in Chapter 18. Condition ii is often easy to check by using the Jacobian criterion, explained in Chapter 16. Thus the following criterion is a powerful tool for proving that a ring is a domain; see Theorem 18.15 and the example following it. We shall deduce the general form of the criterion from the special case given in Theorem 11.2. A related but easier result, characterizing reduced rings, is given in Exercise 11.10.

Theorem 11.5 (Serre's Criterion). *A Noetherian ring R is a direct product of normal domains iff the following two conditions are satisfied:*

i. Every associated prime of a principal ideal generated by a nonzerodivisor in R is of codimension 1; every associated prime of 0 is of codimension 0.

ii. Every localization of R at a codimension-1 prime is a discrete valuation ring; every localization of R at a codimension-0 prime is a field.

Condition ii, regularity in codimension 1, is sometimes called "R1" where the R stands for "Regular." Condition i is usually called "S2" where the S stands, somewhat asymmetrically, for Serre. See Exercise 11.10 for the meaning of the condition "R0 and S1" and the discussion after Theorem 18.15 for a reinterpretation of the conditions S1 and S2.

Proof. If R is a direct product of rings, say $R = R_1 \times \cdots \times R_n$, then any prime of R has the form

$$(*) \qquad R_1 \times \cdots \times R_{i-1} \times Q_i \times R_{i+1} \times \cdots \times R_n$$

for some i and some prime Q_i of R_i. The associated primes of 0 are those of the form $(*)$ with Q_i associated to 0 in R_i. Similarly, the primes associated to a nonzerodivisor

$$(a_1, \ldots, a_n) \qquad a_j \in R_j \text{ a nonzerodivisor for each } j$$

are those of the form $(*)$ with Q_i an associated prime of a_i.

If now each of the R_j is normal, then R satisfies condition i because by Theorem 11.2, each of the R_j does, and it satisfies condition ii because in addition each localization of R at a prime of codimension c is a localization of some R_j at a prime of codimension c.

Conversely, suppose that R satisfies conditions i and ii. We shall first show that R is reduced. If

$$0 = \cap I_j \quad \text{with } I_j \text{ a } P_j\text{-primary ideal}$$

is a minimal primary decomposition of 0, then each P_j is an associated prime of 0 and thus has codimension 0 by condition i. By condition ii, R_P is a field, so that $I_j = P_j$, and R is reduced.

We may now apply Proposition 11.3. Since for each prime P associated to a nonzerodivisor in R the ring R_P is integrally closed, it follows that R itself is integrally closed in $K(R)$. Since R is reduced, $K(R)$ is a reduced zero-dimensional ring, and by Proposition 2.16 $K(R)$ is the product of fields $K_j = (R/P_j)_{P_j}$. Let e_j be the identity element of K_j, so that e_j is an idempotent of $K(R)$ and $e_i e_j = 0$ for $i \neq j$. Since e_j satisfies the integral equation $e_j^2 - e_j = 0$, we must have $e_j \in R$ for each j. It follows that R is the product of the rings $Re_j = R/P_j$. Further, since R is integrally closed in $K(R)$, it now follows that each R/P_j is integrally closed in K_j, so R is a product of normal domains, as required. □

11.3 Invertible Modules

Anyone who has ever looked into a modern paper in algebraic geometry will have seen the phrase "invertible sheaf" or its approximate synonym "line bundle" prominently displayed. The reason is that these notions play a major role in the codimension-one theory of varieties. They will play a major role for us, too. We begin with the definition.

If R is a ring and I is an R-module, then I is **invertible** if I is finitely generated and if for every prime ideal P of R we have $I_P \cong R_P$; that is, I is locally free of rank 1. Of course, it suffices to check this for maximal ideals, since if $P \subset \mathfrak{m}$ are primes then R_P is a further localization of $R_{\mathfrak{m}}$. We write I^* for $\operatorname{Hom}_R(I, R)$, and we make use of the natural map $\mu : I^* \otimes I \to R$ by $\varphi \otimes a \mapsto \varphi(a)$. (We shall see that every invertible module is isomorphic to an ideal, so the name I is not too misleading.)

The simplest invertible ideals are the principal ones: If $I = (x)$, where x is a nonzerodivisor, then I is invertible. It is easy to give an example of a nonprincipal invertible ideal: The ideal $I = (2, 1 + \sqrt{-5}) \subset \mathbf{Z}[\sqrt{-5}]$ is one. (*Proof*: Easy computation shows $2 \in I^2$. If $I = (x)$ for some x, then $2 = ux^2$ for some element u of $\mathbf{Z}[\sqrt{-5}]$. Let $N(a + b\sqrt{-5}) = a^2 + 5b^2$ be the norm. We get $4 = N(2) = N(u)N(x)^2$. Since $N(x)$ is not a unit, $N(x) = \pm 2$. But $a^2 + 5b \neq \pm 2$. Thus I is not principal. To check that I is locally principal at a prime P, first note that if $I \not\subset P$ then $I_P = R_P$. If $I \subset P$ then $2 \in IP$, so by Nakayama's lemma $I_P = (1 + \sqrt{-5})_P$.) In the realm of affine rings, examples are common: For example, all the maximal ideals of $k[x, y]/(y^2 - x^3 + x)$ are invertible but not principal if k is an algebraically closed field of characteristic not 2. (Proof for those who know some geometry: This is the affine ring of a nonsingular curve of genus 1 with just one point at infinity. If a maximal ideal were principal then, because the divisor of a rational function on a complete curve has degree 0, the generator would birationally map the curve to $\mathbf{P}^1$.) Systematic algebraic methods, and some more examples, are given in the exercises.

We shall compare invertible modules with R-submodules of the total quotient ring $K(R)$. These are called **fractional ideals** of R. If I is a finitely generated fractional ideal of R, then choosing a common denominator for the generators of I shows that I is isomorphic to an ordinary ideal of R. If $I \subset K(R)$ is any set, we define $I^{-1} := \{s \in K(R) | sI \subset R\}$.

Theorem 11.6. *Let R be a Noetherian ring.*

a. *If I is an R-module, then I is invertible iff the natural map $\mu : I^* \otimes I \to R$ is an isomorphism.*

b. *Every invertible module is isomorphic to a fractional ideal of R. Every invertible fractional ideal contains a nonzerodivisor of R.*

c. *If $I, J \subset K(R)$ are invertible modules, then the natural maps $I \otimes J \to IJ$, taking $s \otimes t$ to st, and $I^{-1}J \to \operatorname{Hom}_R(I, J)$, taking $t \in I^{-1}J$ to*

$\varphi_t : I \to J$ defined by $\varphi_t(a) = ta$, are isomorphisms. In particular, $I^{-1} \cong I^*$.

d. *If $I \subset K(R)$ is any R-submodule, then I is invertible iff $I^{-1}I = R$.*

Proof.

a. If I is invertible then μ localizes, at any prime P, to the isomorphism $\mu_P : R_P^* \otimes_{R_P} R_P = R_P \otimes_{R_P} R_P \to R_P$. By Corollary 2.9, μ is an isomorphism.

   Conversely, suppose μ is an isomorphism. Suppose that $1 = \mu(\sum_{i=1}^n \varphi_i \otimes a_i)$. It follows that for every prime $P, \mu_P : (I^* \otimes_R I)_P = (I^*)_P \otimes_{R_P} I_P \to R_P$ is an isomorphism. We shall show that $I_P \cong R_P$ and that it is generated by one of the a_i. Some $\varphi_i(a_i)$ must be outside P. Let $v = \varphi_i(a_i)^{-1} \in R_P$ so that $a := va_i$ goes to 1 under $(\varphi_i)_P$. Then $I_P = R_P a \oplus \ker(\varphi_i)_P$, with $R_P a \cong R_P$. Similarly, regarding a as a homomorphism from I_P^* to R_P, we see that $I_P^* = R_P\varphi_i \oplus \ker(a)$ and $R\varphi_i \cong R$. Now,

$$I_P^* \otimes I_P = R_P a \otimes R_P \varphi_i \oplus \ker(\varphi_i)_P \otimes R_P \varphi_i \oplus \cdots .$$

   Since $(\ker \varphi_i) \otimes R_P\varphi_i$ maps to $\varphi_i(\ker \varphi_i) = 0$ under the isomorphism μ, we have $(\ker \varphi_i)_P = 0$, and $(\varphi_i)_P$ is an isomorphism sending a_i to a generator, as claimed. It follows from Corollary 2.9 that I is generated by $a_1, \ldots, a_n$, so I is also finitely generated.

b. Suppose I is invertible. We wish to embed I in $K(R)$. By Exercise 3.14, $K(R)$ is a semilocal ring; its maximal ideals are of the form $PK(R)$, where P is a maximal associated prime of R. For every such P we have $I \otimes K(R)_{PK(R)} = I_P \cong R_P \cong K(R)_{PK(R)}$. Thus by Exercise 4.13, $I \otimes K(R) \cong K(R)$.

   Next we show that the localization map $\varphi : I \to K(R) \otimes I = I[U^{-1}]$, where U is the set of nonzerodivisors on R, is a monomorphism. To prove this, it is enough by Corollary 2.9 to check it locally at a maximal ideal P. The map φ_P is the localization map $\varphi_P : I_P \cong R_P \to K(R) \otimes_P R_P = R_P[U^{-1}]$. The elements of U are nonzerodivisors on R_P, so φ_P is a monomorphism as required. The map $I \to K(R) \otimes I \cong K(R)$ is the desired embedding.

   Suppose now that $I \subset K(R)$ is any finitely generated fractional ideal such that $I \cap R$ consists of zerodivisors. Because I is finitely generated, there is a nonzerodivisor $u \in R$ such that $uI \subset R \cap I \subset R \subset K(R)$. By Corollary 3.2 there is a nonzero element $b \in R$ annihilated by $R \cap I$, and thus by uI. It follows that I is annihilated by ub, and localizing at a prime P containing the annihilator of ub, we see that $I_P \ncong R_P$, so I is not invertible.

c. Suppose that $I, J \subset K(R)$ are invertible. We first show that the natural surjection $I \otimes J \to IJ$ is a monomorphism. It suffices to show that for any prime P of R the map $I_P \otimes_{R_P} J_P \to (IJ)_P \subset K(R)_P$ is a monomorphism. Now $K(R_P)$ is a localization of $K(R)_P$, and it suffices to prove that the composite map to $K(R_P)$ is a monomorphism. Thus we may assume from the outset that R is local.

In this case $I \cong J \cong R$, so I and J are generated as R-modules by some nonzerodivisors s and t of $K(R)$. It follows that st is a nonzerodivisor. The composite map $R \cong R \otimes_R R \cong I \otimes_R J \to IJ = Rst \subset K(R)$ is multiplication by st, a monomorphism as claimed.

Next we show that the natural map $I^{-1}J \to \operatorname{Hom}_R(I, J)$ sending t to φ_t is an isomorphism. By part b we may choose a nonzerodivisor $v \in R \cap I$. If $0 \neq t \in I^{-1}J$, then $tv \neq 0$, so t induces a nonzero element of $\operatorname{Hom}_R(I, J)$, and the map $I^{-1}J \to \operatorname{Hom}_R(I, J)$ is a monomorphism. To show that it is an epimorphism, let $\varphi \in \operatorname{Hom}_R(I, J)$ be arbitrary, and set $\varphi(v) = w$. We claim that $\varphi = \varphi_{w/v}$.

In fact, we claim that if any two homomorphisms $\varphi, \psi : I \to K(R)$ agree on v, then they agree on all of I. It suffices to show that they agree after localization. The element v has the property that its annihilator is 0, and this is preserved by localization, so v corresponds to a nonzerodivisor of R_P under the isomorphism $I_P \cong R_P$. After choosing such an isomorphism, we may regard φ_P and ψ_P as homomorphisms of R_P-modules $R_P \to K(R)_P$ that agree on a nonzerodivisor v of R_P. Thus, $v\varphi(1) = \varphi(v) = \psi(v) = v\psi(1)$, so $\varphi(1) = \psi(1)$ and $\varphi = \psi$.

d. First suppose that $I \subset K(R)$ is an invertible module. By part c the isomorphism $I^* \otimes I \to R$ may be identified with the multiplication map $I^{-1} \otimes I \to R$, so $I^{-1}I = R$.

Finally, suppose $I \subset K(R)$ is an R-submodule with $I^{-1}I = R$. We may localize and suppose that R is local with maximal ideal P, and we must then show that $I \cong R$. By our hypothesis there is an element $v \in I^{-1}$ such that $vI \not\subset P$. It follows that $vI = R$. This implies that v is a nonzerodivisor, so multiplication by v is an isomorphism of I with R.

□

Part d of Theorem 11.6 accounts for the term "invertible" submodule. Since the tensor product is associative, the set of isomorphism classes of invertible R-modules forms a group under the operation $\otimes$: The unit is the isomorphism class of R and by Theorem 11.6a the inverse of the class of I is the class of I^*. This group is called the **Picard group of R**, denoted $\operatorname{Pic}(R)$.

Similarly, the set of invertible submodules of $K(R)$ forms a group under multiplication; the inverse of $I \subset K(R)$ is I^{-1}. This is called the group

of **Cartier divisors**, $C(R)$ (we shall see the origin of the term later). We have:

Corollary 11.7. *Let R be a Noetherian ring.*

a. *The map $C(R) \to \mathrm{Pic}(R)$ sending each invertible submodule of $K(R)$ to its isomorphism class is surjective, and its kernel is isomorphic to the group of units of $K(R)$.*

b. *The group $C(R)$ is generated by the set of invertible ideals of R.*

Proof.

a. To each unit $u \in K(R)$ we associate the "principal divisor" $Ru \subset K(R)$; it is certainly an invertible module, with inverse Ru^{-1}. The map $C(R) \to \mathrm{Pic}(R)$ is surjective by Theorem 11.6b and takes principal divisors to the identity, so it suffices to show that if $I, J \subset K(R)$ are invertible submodules and $\varphi : I \to J$ is an isomorphism, then $I = uJ$ for some unit $u \in K(R)^*$. By part c of Theorem 11.6, $\mathrm{Hom}_R(I, J) = I^{-1}J$, so φ may be realized as multiplication by some element $u \in K(R)$. The inverse map is similarly realized by multiplication by some element $v \in K(R)$. Thus multiplication by uv is the identity on I, and since I must contain a nonzerodivisor of R, we see that $uv = 1$.

b. If $I \subset K(R)$ is an invertible fractional ideal of R, then I^{-1} contains a nonzerodivisor $a \in R$ by Theorem 11.6b. Thus $aI \subset R$, and we may write $I = aI \cdot (a)^{-1}$. □

11.4 Unique Factorization of Codimension-One Ideals

As we have mentioned before, commutative algebra began in the search for an analogue of unique factorization that would hold for the ring of all integers in a given algebraic number field K—that is, for the integral closure of $\mathbf{Z}$ in K. One fruit of this search was the Lasker-Noether theory of primary decomposition, which deals with arbitrary ideals in an arbitrary Noetherian ring, and in which products are replaced by intersections; we treated this in Chapter 3. But long before the work of Lasker and Noether (1905), Dedekind (1871) had described an extension of unique factorization that gave a unique expression of ideals of a ring R of algebraic integers as products of prime ideals. The key fact about R from this point of view turns out to be that its localizations are all factorial. (In Dedekind's case R is normal and one-dimensional, so the localizations are all DVRs.) The

generalization is worthwhile because, as we shall prove later, any regular local ring is factorial. Thus if R is the affine ring of any nonsingular affine variety, then R is locally factorial.

We say that an ideal I in a ring R has **pure codimension 1** if every associated prime ideal of I has codimension 1. We include the case when I has no associated primes at all—that is, when $I = R$.

Theorem 11.8. *Let R be a Noetherian domain, and suppose that for every maximal ideal P of R the ring R_P is factorial.*

a. Let $I \subset R$ be an ideal. I is an invertible module iff I has pure codimension 1.

b. If $I \subset K(R)$ is an invertible fractional ideal, then I is uniquely expressible as a finite product of powers of prime ideals of codimension 1. Thus $C(R)$ is a free abelian group generated by the codimension 1 primes of R.

Proof. Suppose first that $I \subset R$ is an invertible ideal. If we localize at any maximal ideal then I becomes principal, generated by a nonzerodivisor. Since a factorial domain is normal by Proposition 4.10, Theorem 11.2 shows that I is unmixed of codimension 1.

Next suppose that P is a prime ideal of codimension 1. Suppose $\mathfrak{m}$ is a maximal ideal of R. If $P \subset \mathfrak{m}$, then $P_\mathfrak{m} \subset R_\mathfrak{m}$ is principal because $R_\mathfrak{m}$ is factorial. If $P \not\subset \mathfrak{m}$, then $P_\mathfrak{m} = R_\mathfrak{m}$. In either case $P_\mathfrak{m} \cong R_\mathfrak{m}$, so P is invertible.

We shall now show that any ideal I of pure codimension 1 is a finite product of codimension 1 prime ideals. Since a product of invertible ideals is invertible, this will show that I is invertible as well. Arguing by contradiction, let I be an ideal of pure codimension 1, maximal among such ideals that cannot be expressed as the product of codimension 1 prime ideals. Note that $I = R$ is equal to the empty product of prime ideals, so we may assume $I \neq R$. Let P be a codimension 1 prime ideal containing I.

Since P is invertible, we have $P^{-1}P = R$ and thus $P^{-1} \supsetneqq R$. If $P^{-1}I = I$, then P^{-1} would consist of elements integral over R by Corollary 4.6. As R is locally factorial, it is normal by Proposition 4.10, so this is impossible and $P^{-1}I \supsetneqq I$. By our maximality hypothesis, we may write $P^{-1}I = \Pi Q_i$, a finite product of codimension 1 prime ideals, and thus $I = P\Pi Q_i$ is a product of prime ideals after all. By Corollary 11.7b, every invertible fractional ideal may be expressed as a product of powers of codimension 1 prime ideals.

It remains to show that the expression of I as a finite product of powers of distinct codimension 1 prime ideals is unique. Suppose $I = \Pi_{i=1}^m P_i^{d_i} = \Pi_{i=1}^n Q_i^{e_i}$ are two such expressions. Multiplying both sides by any primes that appear to negative powers, we may assume that all the d_i and e_i are greater than 0. We do induction on $d := \sum d_i$. If $d = 0$, then $I = R$ and

$n = 0$ as well. If $d \geq 1$, note that $\Pi P_i \subset Q_1$. Since Q_1 is prime, some P_i must be contained in Q_1. Since P_i and Q_1 are both codimension 1, we must have $P_i = Q_1$. Since Q_1 is invertible, we may "cancel" Q_1 by multiplying both sides by Q_1^{-1}, reducing d and finishing the proof. (For the relation of this to primary decomposition, see Exercise 11.11.) □

A **Dedekind domain** is a Noetherian normal domain of dimension 1. Thus the ring of all integers in an algebraic number field and the affine ring of a nonsingular irreducible algebraic curve are Dedekind domains. Theorem 11.8 is most often stated and applied for Dedekind domains. Here is the statement in this special case.

Corollary 11.9 (Dedekind). *Let R be a Dedekind domain. Every nonzero ideal of R is invertible and may be written uniquely as the product of prime ideals. The same is true for fractional ideals. Thus $C(R)$ is a free abelian group generated by the set of maximal ideals of R.*

If R is a Dedekind domain, then the group $\text{Pic}(R) = C(R)/K(R)^*$ is usually called the **class group** of R. It is an interesting invariant, about which a good deal is known. For example, if R is the ring of integers in a number field, then $\text{Pic}(R)$ is a finite group.

On the other hand, if R is the affine ring of a nonsingular curve over an algebraically closed field, then $\text{Pic}(R)$ is finite iff the curve is rational. For a curve of genus $g > 0$, $\text{Pic}(R)$ may be represented as the **Picard group of the associated complete curve** modulo the subgroup generated by the classes of the ideals of the finitely many points at infinity. Over the complex numbers, for example, the Picard group of a complete curve of genus g is isomorphic as a group to a product of $\mathbf{Z}$ and a torus, the product of $2g$ copies of the circle $\{z \in \mathbf{C} \mid |z| = 1\}$. This is an uncountable group, and an easy argument shows that $\text{Pic}(R)$ is actually a quotient of the torus by a finitely generated group projection onto $\mathbf{Z}$. Thus $\text{Pic}(R)$ is an uncountable, divisible group. If R is the affine ring of a singular curve, then the Picard group of R maps onto that of the normalization of R, which corresponds to a nonsingular curve; the kernel is an interesting invariant of the singularities.

In general, it is known that every abelian group appears as the Picard group of some Dedekind domain, and the sets of generators and relations given by invertible ideals and principal ideals may also be prescribed; see Leedham-Greene [1972].

11.5 Divisors and Multiplicities

If a, b are elements of a ring, then b divides a iff $a \in (b)$. In general, an ideal can be regarded as something by which an element might be divisible—a "divisor." Because of the unique factorization into prime ideals in a Dedekind domain, nonzero ideals there correspond to finite sets of codimension 1 prime ideals, each with multiplicity. The term *divisor* was transferred to such sets and stuck there. We define a **divisor** (or **Weil divisor**) in any ring R to be an element of the free abelian group $\mathrm{Div}(R)$ whose generators are the codimension 1 prime ideals of R. That is, a Weil divisor of R is a formal linear combination of codimension 1 prime ideals in R, with integer coefficients.

On the other hand, the natural analogue, for an arbitrary ring R, of the set of divisors in a Dedekind domain is in many respects the set of invertible ideals of R, now called **Cartier divisors.** (Both the names Cartier divisor and Weil divisor seem to have been coined by Mumford [1966].) In general, these two sets are very different, but there is a natural homomorphism from the group of Cartier divisors to the group of Weil divisors, which we shall describe.

We shall exploit the following elementary fact: If R is a one-dimensional ring and $a \in R$ is any nonzerodivisor, then $R/(a)$ is zero-dimensional, and thus of finite length. We shall see that the map $R - \{0\} \to \mathbf{Z}$ defined by $a \mapsto \operatorname{length} R/(a)$ extends to a homomorphism from $K(R)^*$ to $\mathbf{Z}$. (In general it is *not* a valuation; see Exercise 11.6.) This homomorphism even extends to the group of invertible ideals.

Theorem 11.10. *For any Noetherian ring R, there is a map $\varphi : C(R) \to \mathrm{Div}(R)$ sending an invertible ideal $I \subset R$ to*

$$\varphi(I) = \sum_{P \subset R \text{ codim 1 prime}} \operatorname{length}(R_P/I_P) \cdot P \in \mathrm{Div}(R).$$

If $\dim R = 1$, *then there is a map $C(R) \to \mathbf{Z}$ sending an invertible ideal I to* $\operatorname{length} R/I$.

Proof. We begin with a general remark: Suppose that G, H are abelian groups and that $\mathfrak{s} \subset G$ is a subset that generates G. To define a homomorphism $\varphi : G \to H$, it suffices to give φ on $\mathfrak{s}$, and check that φ respects products of elements of $\mathfrak{s}$ in the sense that if $\Pi a_i = \Pi b_j$ with a_i and b_j in $\mathfrak{s}$, then $\Pi \varphi a_i = \Pi \varphi b_j$. That is, one need not worry about the inverses of elements of $\mathfrak{s}$. To extend φ from $\mathfrak{s}$ to all of G, suppose $x = \Pi s_i \Pi t_j^{-1} \in G$ with $s_i, t_j \in \mathfrak{s}$. We would like to set φx equal to $\Pi \varphi s_i \Pi(\varphi t_j)^{-1}$. If $x = \Pi s_i' \Pi t_j'^{-1}$ is another expression of x in terms of elements in $\mathfrak{s}$, then $\Pi s_i \Pi t_j' = \Pi s_i' \Pi t_j$. Since we have assumed that φ is well defined on products of generators, $\Pi \varphi s_i \Pi \varphi t_j' = \Pi \varphi s_i' \Pi \varphi t_j$ and $\Pi \varphi s_i \Pi(\varphi t_j)^{-1} = \Pi \varphi s_i' \Pi(\varphi t_j')^{-1}$. Thus, we may

define φx as above, obtaining a map of sets $\varphi : G \to H$. It follows at once that φ is a group homomorphism.

We now turn to the definition of $\varphi : C(R) \to \mathrm{Div}(R)$. The set $\mathfrak{s}$ of all invertible ideals $I \subset R$ generates $C(R)$, and by the beginning remark it suffices to show that the formula above defines $\varphi(I)$ for $I \in \mathfrak{s}$ and prove that φ respects products. Suppose $I \subset R$ is an invertible ideal and P is a codimension-1 prime of R. The localization R_P is one-dimensional, and I_P contains a nonzerodivisor, so R_P/I_P is zero-dimensional and length $(R_P/I_P) < \infty$. If $I \not\subset P$, then the length is 0. If $I \subset P$, then since I contains a nonzerodivisor, P must be one of the finitely many minimal primes of I. Thus the sum defining $\varphi(I)$ is finite, and φ is well defined on the generators $\mathfrak{s}$.

To show that φ respects products, suppose that $I = \Pi I_j$ where each I_j is an invertible ideal contained in R. We must show that for every codimension-1 prime P, $\mathrm{length}(R_P/I_P) = \sum_j \mathrm{length}(R_P/I_{jP})$. To simplify the notation we may suppose that R itself is local and one-dimensional.

Since I_j is invertible, it is a principal ideal generated by some nonzerodivisor, $a_j \in R$. We must show that length $R/(\Pi_j a_j) = \sum_j$ length $R/(a_j)$. Consider the filtration

$$R \supset (a_1) \supset (a_1 a_2) \supset \cdots \supset (\Pi a_j).$$

To prove the length equality it suffices to show that $(\Pi_{j<i} a_j)/(\Pi_{j\le i} a_j) \cong R/(a_i)$. Since each a_j is a nonzerodivisor, multiplication by $b := \Pi_{j<i} a_j$ induces an isomorphism $R \to (\Pi_{j<i} a_j)$. If for some $a \in R$ we have $ba \in (a_i b)$, say $ba = a_i br$, then since b is a nonzerodivisor we must have $a = a_i r \in (a_i)$. Thus the preimage of $(\Pi_{j\le i} a_j)$ is (a_i), so multiplication by b induces the desired isomorphism.

For the second statement of the theorem, suppose that R is one-dimensional (but not necessarily local). If $I \subset R$ is invertible, then I contains a nonzerodivisor so that $\dim R/I = 0$ and length $R/I < \infty$. The minimal primes of R/I are the codimension-1 primes of R that contain I. Thus by Theorem 2.13,

$$\mathrm{length}\, R/I = \sum_{P \text{ a codim-1 prime containing } I} \mathrm{length}(R_P/I_P).$$

That is, the length is obtained by composing $\varphi : C(R) \to \mathrm{Div}(R)$ with the homomorphism $\mathrm{Div}(R) \to \mathbf{Z}$ that takes $\sum_{P \text{a codim}-1 \text{ prime}} n_P P$ to $\sum n_P$. □

See Exercise 11.13 for some refinements of the length map.

If $a \in K(R)^*$ then we say that the divisor $\varphi((a))$ is a principal divisor; the group of divisors modulo the principal divisors is called $\mathrm{Chow}(R)$, the **codimension-1 Chow group** of R. (See Exercise 11.12 for the relation to the usual definition of the Chow groups.) Thus the map φ induces a map $\psi : \mathrm{Pic}(R) \to \mathrm{Chow}(R)$. Theorem 11.8 shows that this map is an isomorphism

when R is locally factorial, but in general ψ is neither surjective (there may be codimension-1 primes that are not invertible) nor injective (it may not be possible to distinguish invertible ideals by the "numerical" information of the associated divisor). For example, see Exercise 11.17.

Proposition 11.11. *If R is a normal Noetherian ring, then the maps $\varphi : C(R) \to \mathrm{Div}(R)$ and $\psi : \mathrm{Pic}(R) \to \mathrm{Chow}(R)$ are injective.*

Proof. Consider the diagram

$$\begin{array}{ccccccccc} 0 & \to & K(R)^* & \to & C(R) & \to & \mathrm{Pic}(R) & \to & 0 \\ & & \| & & {\scriptstyle\varphi}\downarrow & & {\scriptstyle\psi}\downarrow & & \\ & & K(R)^* & \to & \mathrm{Div}(R) & \to & \mathrm{Chow}(R) & \to & 0. \end{array}$$

From the diagram it follows that if φ is a monomorphism then the map $K(R)^* \to \mathrm{Div}(R)$ is a monomorphism and (for example, by the "snake lemma," Exercise A3.10) ψ is a monomorphism. Thus it will suffice to show that φ is a monomorphism. Since any invertible fractional ideal may be written as one invertible ideal of R times the inverse of another, it suffices to show that two invertible ideals $I, J \subset R$ that have the same divisor are equal.

By symmetry it will suffice to prove that $I \subset J$. By Corollary 3.5, applied with $M = R/J$, it is enough to show that $I_P \subset J_P$ for any associated prime P of J. Now P_P is an associated prime of J_P, which is a principal ideal generated by a nonzerodivisor, and R_P is normal, so by Serre's criterion (Theorem 11.5), P has codimension 1 and R_P is a discrete valuation ring. Since I and J have the same divisor, we have length $R_P/I_P =$ length R_P/J_P. Since the ideals of R_P are totally ordered it follows that $I_P = J_P$, and this proves what is required. □

11.6 Multiplicity of Principal Ideals

Suppose that R is a one-dimensional ring. In Lemma 11.10 we saw that the length of $R/(a)$ gives a homomorphism from $K(R)^\times$ to $\mathbf{Z}$. The basic result of this section expresses the length of M/aM for certain R-modules M in terms of the length of $R/(a) = R/aR$ and an invariant of M. We shall apply this to prove Krull's theorem that the integral closure of a one-dimensional Noetherian ring is Noetherian, in the strong form due to Akizuki [1935].

For simplicity, and because it suffices for the application, we shall assume here that R is a domain. Let $K = K(R)$ be the quotient field. An R-module M is said to be **torsion-free** if every nonzero element of R is a nonzerodivisor on M, or equivalently if the localization map $M \to K \otimes_R M$ is a monomorphism. Note that $K \otimes_R M$ is a vector space over the field K. Set rank $M := \dim_K K \otimes M$.

Lemma 11.12. *Let R be a one-dimensional Noetherian domain. If M is a torsion-free R-module, then*

$$\operatorname{length}(M/xM) \le \operatorname{rank}(M) \operatorname{length}(R/(x)),$$

with equality if M is finitely generated as an R-module.

Note that the case $M = K$ shows that we really may have length $M/xM < \operatorname{rank}(M) \operatorname{length} R/xR$ if M is not finitely generated. We have not included the case of invertible ideals because the application does not require it; see Exercise 11.14.

Proof. Set $r = \operatorname{rank}(M)$. If M is finitely generated then $r < \infty$; thus if $r = \infty$ there is nothing to prove, and we may assume that $r < \infty$.

Assume that M is finitely generated. We may choose a K-basis $m_1, \ldots, m_r$ of $K \otimes_R M$ inside M: To do this, first choose any basis. Each basis element can be written as a K-linear combination of elements of M. Multiplying each basis element by a suitable element of R (which is a unit of K) we get a new basis coming from M.

The map $\alpha : R^r \to M$ sending the elements of a basis of R^r to the m_i becomes an isomorphism after tensoring with K. Thus the kernel is 0, and we may think of R^r as a submodule of M. Let $N = M/R^r$.

We shall use the elementary properties of Tor introduced in section 6.2. Tensoring the exact sequence $0 \to R^r \to M \to N \to 0$ with R/xR we get an exact sequence

$$\operatorname{Tor}_1^R(M, R/xR) \to \operatorname{Tor}_1^R(N, R/xR) \to$$
$$R^r/xR^r \xrightarrow{\bar{\alpha}} M/xM \to N/xN \to 0,$$

where we have written $\bar{\alpha}$ for the induced map $R^r/xR^r \to M/xM$.

By the computation done at the end of section 6.2 we have $\operatorname{Tor}_1^R(M, R/xR) = (0:_M x)$ the set of elements of M killed by x. This is 0 since M is torsion-free. Similarly, we get $\operatorname{Tor}_1^R(N, R/xR) = (0:_N x)$. Thus $\ker \bar{\alpha} = (0:_N x)$. Putting this together we get an exact sequence

$$(*) \qquad 0 \to (0:_N x) \to R^r/xR^r \to M/xM \to N/xN \to 0.$$

To analyze $(0:_N x)$ we use the exact sequence that comes from multiplication by x on N:

$$0 \to (0:_N x) \to N \xrightarrow{x} N \to N/xN \to 0.$$

Since $K \otimes \alpha$ is an isomorphism, and tensoring with K is exact, we have $K \otimes_R N = 0$. Since $K \otimes_R N$ is the localization of N inverting the set of nonzero elements of R, each element of N is annihilated by a nonzero element of R. Since M is finitely generated, so is N, and we see that some nonzero element $f \in R$ annihilates all of N. Since R is 1-dimensional, R/fR

is 0-dimensional, and thus of finite length. Since N is a finitely generated R/fR-module, N has finite length.

Since the alternating sum of the lengths of the terms of any exact sequence of modules of finite length is 0, we get

$$\begin{aligned}\operatorname{length}(0:_N x) &\\ &= \operatorname{length} N - \operatorname{length} N + \operatorname{length} N/xN\\ &= \operatorname{length} N/xN.\end{aligned}$$

Using the same principle with the sequence $*$) we get

$$\begin{aligned}\operatorname{length} M/xM &\\ &= \operatorname{length} R^r/xR^r + \operatorname{length} N/xN - \operatorname{length}(0:_N x)\\ &= \operatorname{length} R^r/xR^r\\ &= r \operatorname{length} R/xR,\end{aligned}$$

proving the equality in the Lemma in case M is finitely generated.

We now deduce the inequality of the Lemma in the general case from the finitely generated case. No longer supposing that M is finitely generated, suppose that contrary to the assertion in the Lemma we have length $M/xM > r$ length R/xR. We may choose a finitely generated submodule $M' \subset M$ whose image N' in M/xM has length $> r$ length R/xR. But then

$$\begin{aligned}\operatorname{length} M'/xM' &\\ &\geq \operatorname{length} N'\\ &> r \operatorname{length} R/xR\\ &\geq \operatorname{rank}(M') \operatorname{length} R/xR,\end{aligned}$$

contradicting the equality that we have proved in the finitely generated case. □

As an application we prove that the integral closure of a one-dimensional Noetherian domain R is Noetherian. We shall prove it in a particularly strong form: We show that any ring contained between R and a finite extension of the quotient field $K(R)$ is Noetherian; this is the Krull-Akizuki theorem. In dimension two, Mori and Nagata prove that the integral closure of Noetherian domain is Noetherian (see Nagata [1962, Theorem 33.12]), though in this case there may be subrings of the integral closure containing R but not Noetherian. In dimension at least 3 there are unfortunately Noetherian domains whose integral closures are not Noetherian. For examples of all these things, see Nagata [1962, Appendix, Examples 3–5]. In Chapter 13 we shall show that if R is an affine domain of any dimension, then the integral closure of R (in any finite extension of $K(R)$) is actually a finitely generated R-module—something that is not true even for one-dimensional Noetherian domains.

The most interesting case in which to apply the lemma is when M is the integral closure of R in its quotient field:

Theorem 11.13 (Krull-Akizuki Theorem). *If R is a one-dimensional Noetherian domain with quotient field K, and L is a finite extension field of K, then any subring S of L that contains R is Noetherian and of dimension ≤ 1, and has only finitely many ideals containing a given nonzero ideal of R. In particular, the integral closure of R in L is Noetherian.*

Proof. Let $0 \neq J \subset S$ be an ideal. S is algebraic over R, so J intersects R nontrivially by Lemma 4.16; let $0 \neq a \in R$ be an element of $J \cap R$. All the assertions of the theorem follow immediately if we show that J/aS is a module of finite length. Since $J/aS \subset S/aS$, it suffices to show that S/aS is a module of finite length. Since $K(R) \otimes S \subset L, S$ is a torsion-free R-module of finite rank, so we are done by Lemma 11.12. □

11.7 Exercises

Valuation Rings

Exercise 11.1: A **valuation ring** is a domain R such that the quotient field of R has a valuation ν on it such that R is the valuation ring of ν in the sense of the text. (The name is standard, despite the fact that the term "valuation domain" would seem more sensible).

a.* Show that R is a valuation ring iff R is a domain such that for all $x \in K(R)$, either x or x^{-1} is in R.

b. Show that any valuation ring is integrally closed in its quotient field.

Exercise 11.2 (Existence of valuation rings): Let R be any domain, and let $P \subset R$ be a prime ideal. Show by Zorn's lemma that there exists a subring $R' \subset K(R)$ containing R and maximal among subrings such that $PR' \neq R'$. Show that

a.* R' is local, integrally closed in $K(R), R_P \subset R'$, and if $\mathfrak{m} \subset R'$ is the maximal ideal, then $\mathfrak{m} \cap R = P$.

b. For any domain S and any element $s \in K(S)$, write $S[s]$ for the subring of $K(S)$ generated by S and s. Show that $sS[s] = S[s]$ iff s^{-1} is integral over S. Use this remark to prove the following two facts:

c. R' is a valuation ring.

d. The integral closure of R is the intersection of the valuation rings containing R.

Exercise 11.3: Show that a valuation ring is Noetherian iff it is a DVR.

Exercise 11.4: Let G be any ordered abelian group. (That is, G is an abelian group that as a set is totally ordered by a relation $<$; and if $a, b, c \in G$ and $a < b$, then $a + c < b + c$.) Let k be a field, and let R be the vector space over k with basis $\{x^a | a \in G, a \geq 0\}$; here x^a is simply a symbol, not "x raised to the power a." We define multiplication of basis elements by $x^a x^b = x^{a+b}$. We extend this to a product operation on R by linearity. Show that this operation makes R into a valuation ring with value group G and valuation

$$\nu(\sum r_a x^a) = \min\{a | r_a \neq 0\},$$

where the r_a are elements of k.

Exercise 11.5: Let G be the group $\mathbf{Z} \oplus \mathbf{Z}$, ordered lexicographically; that is, $(n, m) < (n', m')$ iff $n < n'$ or $n = n'$ and $m < m'$. Let R be the valuation ring constructed from G as in Exercise 11.4. Compute the Krull dimension of R.

Exercise 11.6: Show that the function $R - \{0\} \to \mathbf{Z}$ given by $a \mapsto$ length $R/(a)$ is not a valuation in the case $R = k[x]$, with k a field, though it is a valuation on $R = k[x]_{(x)}$.

The Grothendieck Ring

Exercise 11.7 (Projective Modules and the Grothendieck Ring):

a. Show that any invertible R-module is a direct summand of a free R-module. Summands of free modules are called **projective** modules; see Exercise 4.11.

b. Generalizing Theorem 11.6c, show that for any finitely generated projective module M, and any module N, we have $\mathrm{Hom}(M, N) \cong M^* \otimes_R N$.

c. Let $K_0(R)$ be the free group on the isomorphism classes of projective R-modules modulo the relations $[P]+[Q] = [P \oplus Q]$, where $[P]$ denotes the class of a projective module P in $K_0(R)$. Show that if P and Q are projective modules, then $[P] = [Q]$ iff $P \oplus R^n \cong Q \oplus R^n$ for some number n. If P and Q are invertible and $[P] = [Q]$, show that $P \cong Q$. Show that the operation $\otimes_R$ makes $K_0(R)$ into a ring whose group of units is $\mathrm{Pic}(R)$. See also Section 19.4.

Exercise 11.8: If I is an invertible module over a domain R, show that $\mathrm{End}(I) = R$, with elements of R acting by multiplication on I; in particular, the group of automorphisms of I is the group of units of R.

Exercise 11.9 (Rational Maps of Normal Varieties Are Regular in Codimension 1):* (For those who know some algebraic geometry). Deduce from Serre's criterion (Theorem 11.2 is enough) the following geometric statement: A rational map from a normal variety over an algebraically closed field to a projective variety is a morphism on the complement of a set of codimension two. (The algebraically closed assumption is unnecessary if one works with schemes.)

Exercise 11.10 (Characterization of reduced rings):* The following is an analogue of Serre's criterion: Prove that a Noetherian ring R is reduced (that is, has no nilpotents) iff it satisfies

R0: The localization of R at each prime of codimension 0 is regular.
S1: All primes associated to zero have codimension 0.

See the discussion after Theorem 18.15 for a reinterpretation.

Exercise 11.11: Let I be an ideal of pure codimension 1 in a locally factorial ring R (for example, a Dedekind domain). Prove:

a. I is a power of a prime ideal P iff I is P-primary.

b. If $I = \Pi P_i^{n_i}$ with P_i prime and $P_i \neq P_j$ for $i \neq j$, then $I = \cap P_i^{n_i}$ is the primary decomposition of I.

Compare this result with the result of Proposition 3.11.

Exercise 11.12 (Chow groups):* Let R be a Noetherian ring. Let $Z_i(R)$ be the free abelian group on the set of i-dimensional primes of R. Let $A_i(R)$ be $Z_i(R)$ modulo the subgroup generated by all elements that can be written as principal divisors modulo some $(i+1)$-dimensional prime of R. Suppose that in R, as in any affine domain, we have dim P + codim P = dim R for every prime P. Show that $A_1(R)$ is a quotient of the group that we have defined as Chow(R), the codimension-1 Chow group. (The groups $A_i(R)$ are the usual Chow groups; the definition in the text, though natural, is somewhat nonstandard. See Fulton [1984] for more of this story.)

Exercise 11.13: As we have seen, in the one-dimensional case the map sending an invertible ideal I to length R/I is obtained by composing the map φ with the map $\text{Div}(R) \to \mathbf{Z}$ sending each prime ideal to 1. But the different prime ideals may have very different "sizes," and there may be a more interesting map, reflecting these sizes. For example:

a. Suppose that R is a one-dimensional affine domain over a field k. Then for any codimension-1 prime ideal P the field R/P is a finite extension of k. We may compose φ with the map $\text{Div}(R) \to \mathbf{Z}$ sending P to $\dim_k R/P$. Show that the resulting homomorphism $C(R) \to \mathbf{Z}$ sends each invertible ideal $I \subset R$ to $\dim_k R/I$.

b. Suppose that $\mathbf{Z} \subset R$ and R is a finitely generated $\mathbf{Z}$-module. If we compose φ with the map $\mathrm{Div}(R) \to \mathbf{Z}$ sending P to the cardinality of the finite set R/P (see Exercise 4.26), show that the resulting homomorphism $C(R) \to \mathbf{Z}$ sends each invertible ideal $I \subset R$ to card R/I.

Exercise 11.14: State and prove a version of Lemma 11.12 replacing x with an invertible ideal.

Exercise 11.15 (A Method for Constructing Projective Modules): The following is adapted from Milnor's beautiful book [1971, pp. 19–24]: Suppose that we have a commutative diagram of rings and ring homomorphisms

$$\begin{array}{ccc} R & \xrightarrow{\alpha_1} & R_1 \\ \alpha_2 \downarrow & & \downarrow \beta_1 \\ R_2 & \xrightarrow[\beta_2]{} & S \end{array}$$

that is a **fiber square** in the sense that the map $(\alpha_1, \alpha_2) : R \to R_1 \times R_2$ identifies R with the set $\{(a_1, a_2) \in R_1 \times R_2 | \beta_1(a_1) = \beta_2(a_2)\}$, and suppose, moreover, that β_1 is surjective. If P_i is a module over R_i for $i = 1, 2$, and $\varphi : P_1 \otimes_{R_1} S \to P_2 \otimes_{R_2} S$ is an isomorphism of S-modules, then we set

$$M(P_1, P_2, \varphi) := \{(p_1, p_2) \in P_1 \times P_2 | \varphi(p_1 \otimes 1) = p_2 \otimes 1\}.$$

Show that $M(P_1, P_2, \varphi)$ is an R-module in a natural way. Now show:

a. If P_1 and P_2 are free modules, and if with respect to some choice of bases $\{e_i^1\}$ of P_1 and $\{e_i^2\}$ of P_2 the invertible matrix over S that defines φ comes from an invertible matrix φ_1 over R_1, then $M(P_1, P_2, \varphi)$ is free, with basis $\{(\varphi_1^{-1}(e_i^1), e_i^2)\}$. In this case $M(P_1, P_2, \varphi) \otimes R_i \cong P_i$.

b. Now suppose that P_1 and P_2 are free modules and that φ is an arbitrary isomorphism. Use the identity

$$\begin{pmatrix} \varphi & 0 \\ 0 & \varphi^{-1} \end{pmatrix} = \begin{pmatrix} 1 & \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\varphi^{-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

to show that $\varphi \oplus \varphi^{-1}$ lifts to an isomorphism over R_1, and thus that $M(P_1, P_2, \varphi)$ is a projective module.

c. Show that if $M_i = P_i \oplus Q_i$ and $\varphi = \varphi' \oplus \varphi''$ where $\varphi' : S \otimes P_1 \to S \otimes P_2$ and $\varphi'' : S \otimes Q_1 \to S \otimes Q_2$ are isomorphisms, then

$$M(M_1, M_2, \varphi) = M(P_1, P_2, \varphi') \oplus M(Q_1, Q_2, \varphi'').$$

d. If P_1 and P_2 are merely supposed to be projective modules, then there are modules Q_i over R_i such that $P_i \oplus Q_i = R_i^{s_i}$ is free over R_i. Show that if we set $Q_1' = Q_1 \oplus R_1^{s_2}$ and $Q_2' = Q_2 \oplus R_2^{s_1}$, then $S \otimes Q_1' \cong S \otimes Q_2'$. Conclude that $M(P_1, P_2, \varphi)$ is projective. Moreover, if P_1 and P_2 are finitely generated (as modules over R_1 and R_2, respectively), then $M(P_1, P_2, \varphi)$ is finitely generated over R.

e. If P_1 and P_2 are projective modules, then use the last part of a to show that $R_i \otimes P \cong P_i$.

f. Suppose that the P_i are projective over R_i. Show that $M(P_1, P_2, \varphi) \cong M(P_1', P_2', \varphi')$ as R-modules iff there are isomorphisms $\psi_i P_i \to P_i'$ such that $\varphi = S \otimes \psi_2^{-1} \circ \varphi' \circ S \otimes \psi_1$.

g. Show that if $M_i = P_i \otimes_{R_i} Q_i$ and $\varphi = \varphi' \otimes_S \varphi''$ where $\varphi' : S \otimes P_1 \to S \otimes P_2$ and $\varphi'' : S \otimes Q_1 \to S \otimes Q_2$ are isomorphisms, and if P_i and Q_i are projective modules over R_i for $i = 1, 2$, then

$$M(M_1, M_2, \varphi) = M(P_1, P_2, \varphi') \otimes M(Q_1, Q_2, \varphi'').$$

h. The map sending an invertible module I over R to the invertible module $R_i \otimes_R I$ over R_i is a group homomorphism $\delta_i : \text{Pic}(R) \to \text{Pic}(R_i)$. Use part f to show that the kernel of $(\delta_1, \delta_2) : \text{Pic}(R) \to \text{Pic}(R_1) \oplus \text{Pic}(R_2)$ is the group of units of S modulo the images of the groups of units of R_1 and of R_2. That is, writing $U(R)$ for the group of units of R, we have the beginning of a "Mayer-Vietoris sequence"

$$0 \to U(R) \to U(R_1) \oplus U(R_2) \to U(S) \to \text{Pic}(R) \to \text{Pic}(R_1) \oplus \text{Pic}(R_2).$$

Exercise 11.16 (The Conductor Square): Let R be a Noetherian domain, and let R_1 be the normalization of R. Suppose that R_1 is a finitely generated R-module, and let $\mathfrak{c}$ be the annihilator in R of the R-module R_1/R. The ideal $\mathfrak{c}$ is called the **conductor** of R_1 into R, or simply the **conductor** of R.

a. Show that $\mathfrak{c} \subset R \subset R_1$ and $\mathfrak{c}$ is also an ideal of R_1.

b. Show that the natural diagram

$$\begin{array}{ccc} & \alpha_1 & \\ R & \to & R_1 \\ \alpha_2 \downarrow & & \downarrow \beta_1 \\ R/\mathfrak{c} & \to & R_1/\mathfrak{c} \\ & \beta_2 & \end{array}$$

where α_1 and β_2 are the natural inclusions and α_2, β_1 are the natural projections, is a fiber square in the sense of Exercise 11.15.

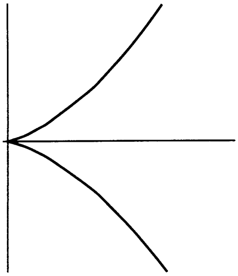


FIGURE 11.1.

c. Suppose that k is a field and $R = k[t^2, t^3] \cong k[x,y]/(y^2 - x^3)$, the affine ring of the rational curve with a cusp, shown in Figure 11.1.

Show that $R_1 = k[t], R/\mathfrak{c} = k, R_1/\mathfrak{c} = k[x]/(x^2)$. Show that the group of units $U(R_1/\mathfrak{c}) = U(k) \oplus k_+$, where k_+ is the additive group of k, embedded in $U(k[x]/(x^2))$ as the set $\{1 + ax | a \in k\}$. Conclude that $\mathrm{Pic}(R) = k_+$.

d. Suppose that k is a field. Let $R = k[t^2 - 1, t^3 - t] \cong k[x,y]/(y^2 - x^2(x+1))$, be the affine ring of the rational curve with a node, shown in Figure 11.2.

Show that $R_1 = k[t]$, $R/\mathfrak{c} = k$, $R_1/\mathfrak{c} = k \times k$. Show that the group of units $U(R_1/\mathfrak{c}) = k^\times \oplus k^\times$. Conclude that $\mathrm{Pic}(R) = k^\times$. What is $\mathrm{Div}(R)/K(R)^\times$?

Exercise 11.17: Let k be a field, and let $R = k[x,y]/(y^2 - x^3)$. Show that the codimension-1 Chow group $\mathrm{Div}(R)/K(R)^\times = 0$; thus $\varphi : \mathrm{Pic}(R) \to \mathrm{Div}(R)/K(R)^\times$ is not a monomorphism.

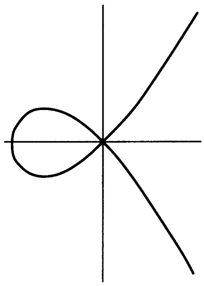


FIGURE 11.2.

12

Dimension and Hilbert-Samuel Polynomials

Note: Throughout this section, all rings considered will be Noetherian.

In this section we present a characterization of dimension that yields other important invariants and that is well suited to computation using techniques to be developed in Chapter 15. Throughout we shall write R for a local ring with maximal ideal $\mathfrak{m}$. All modules will be finitely generated R-modules unless otherwise stated.

If M is a finitely generated R-module, then the Hilbert function $H_M(n)$ is defined by

$$H_M(n) := \dim_{R/\mathfrak{m}} \mathfrak{m}^n M/\mathfrak{m}^{n+1}M.$$

Hilbert himself originally worked with graded rings $R = R_0 \oplus R_1 \oplus \cdots$ generated by R_1 over a field R_0. The local case is not really more general, since M and $gr_{\mathfrak{m}}M$ clearly have the same Hilbert function. In particular, it follows from the graded case (Chapter 1) that $H_M(n)$ agrees with some polynomial function $P_M(n)$ for large values of n. We shall study this polynomial. Our principal goal is to prove the following theorem.

Theorem 12.1. *If R is a local ring, then* $\dim R = 1 + \deg P_R$.

The polynomial that is identically 0 is taken to have degree -1 so that a special case of the theorem says something we already know: $\dim R = 0$ iff $\mathfrak{m}$ is nilpotent.

Samuel [1951] showed that in the context of intersection theory it is useful to study more general polynomial functions associated to a local ring and an ideal primary to the maximal ideal. It turns out that the more general

setting is technically more convenient as well. We shall adopt this context, which we now introduce.

12.1 Hilbert-Samuel Functions

First, we replace the ring R by an arbitrary finitely generated R-module M; we shall show that $\dim M = 1 + \deg P_M$. Generalizing further, we replace the maximal ideal $\mathfrak{m}$ by any parameter ideal $\mathfrak{q}$ for M. We define the **Hilbert-Samuel function** of M with respect to a parameter ideal $\mathfrak{q}$ of M to be

$$H_{\mathfrak{q},M}(n) := \operatorname{length} \mathfrak{q}^n M/\mathfrak{q}^{n+1} M.$$

The given length (that is, the length of a composition series) is finite because $\mathfrak{q}^n M/\mathfrak{q}^{n+1}M$ is a finitely generated module over the Artinian ring $R/(\mathfrak{q} + \text{annihilator } (M))$. We shall show that $H_{\mathfrak{q},M}(n)$ agrees with a polynomial $P_{\mathfrak{q},M}(n)$ and that $\dim M = 1 + \deg P_{\mathfrak{q},M}$.

We shall repeatedly use the elementary fact that if $H(n)$ is any function on the natural numbers whose first difference $G(n) := H(n+1) - H(n)$ is a polynomial of degree d, then $H(n)$ is a polynomial of degree $d+1$ (see Lemma 1.12 for an elaboration).

First we prove:

Proposition 12.2. *Let R be a Noetherian ring and let M be a finitely generated R-module. Suppose either that R is local and $\mathfrak{q} \subset R$ is a parameter ideal for M, or else that $R = R_0 \oplus R_1 \oplus \ldots$ is a graded ring with R_0 Artinian and R generated as an R_0-algebra by R_1. In the latter case set $\mathfrak{q} = (R_1) = R_1 \oplus R_2 \oplus \ldots$. There is a polynomial $P_{\mathfrak{q},M}(n)$ whose degree is $<$ the number of generators of $\mathfrak{q}$ such that for sufficiently large n we have $P_{\mathfrak{q},M}(n) = H_{\mathfrak{q},M}(n)$.*

Proof. If R is local we may first factor out $\operatorname{ann}(M)$ and then pass to $\operatorname{gr}_{\mathfrak{q}} R$ and $\operatorname{gr}_{\mathfrak{q}} M$, so it suffices to do the case where R is graded.

Suppose $x_1, \ldots, x_r \in R_1$ generate $\mathfrak{q}$. We do induction on r, the case $r = 0$ being trivial.

From the exact sequence

$$0 \to (0 :_M x_1) \to M \xrightarrow{x_1} M(1) \to (M/x_1 M)(1) \to 0$$

(where $(0 :_M x_1)$ denotes $\{m \in M | x_1 m = 0\}$ and $M(1)$ denotes the same module as M but with grading shifted so that $M(1)_n = M_{n+1}$), we see that

$$H_{\mathfrak{q},M}(n+1) - H_{\mathfrak{q},M}(n) = H_{\mathfrak{q},M/x_1,M}(n+1) - H_{\mathfrak{q},(0:_M x_1)}(n).$$

The right-hand side is the difference of two Hilbert-Samuel functions of modules over $R/(x_1)$. By induction, they agree for large n with polynomials of degree $< r-1$. The left-hand side is the first difference of the function

$H_{\mathfrak{q},M}(n)$. Thus $H_{\mathfrak{q},M}(n)$ agrees for large n with a polynomial of degree $< r$. □

It is convenient to introduce another function,

$$L_{\mathfrak{q},M}(n) := \text{length}\, M/\mathfrak{q}^n M.$$

From the exact sequences

$$0 \to \mathfrak{q}^n M/\mathfrak{q}^{n+1}M \to M/\mathfrak{q}^{n+1}M \to M/\mathfrak{q}^n M \to 0$$

We see that the first difference function of $L_{\mathfrak{q},M}$ is

$$L_{\mathfrak{q},M}(n+1) - L_{\mathfrak{q},M}(n) = H_{\mathfrak{q},M}(n).$$

It follows from the proposition that $L_{\mathfrak{q},M}(n)$ agrees, for large n, with a polynomial whose degree is 1 + degree $P_{\mathfrak{q},M}$.

In a short exact sequence of graded modules, the Hilbert function of the middle module is the sum of the Hilbert functions of the modules on the ends. In the local case treated here things are not so simple, but thanks to the Artin-Rees lemma, additivity does not fail too badly

Lemma 12.3 (Additivity). *Let R be a local ring. If*

$$0 \to M' \to M \to M'' \to 0$$

is an exact sequence of finitely generated R-modules, and if $\mathfrak{q}$ is a parameter ideal for M, then

$$P_{\mathfrak{q},M} = P_{\mathfrak{q},M'} + P_{\mathfrak{q},M''} - F,$$

where F is a polynomial with positive leading term and whose degree is strictly less than that of $P_{\mathfrak{q},M'}$.

See Exercise 12.10 for a more precise result.

Proof. Since $P_{\mathfrak{q},M}(n) = L_{\mathfrak{q},M}(n+1) - L_{\mathfrak{q},M}(n)$ for large n, it suffices to prove the corresponding result with P replaced by L. From the exact sequence

$$0 \to (M' \cap \mathfrak{q}^n M)/\mathfrak{q}^n M' \to M'/\mathfrak{q}^n M' \to M/\mathfrak{q}^n M \to M''/\mathfrak{q}^n M'' \to 0$$

We see that

$$L_{\mathfrak{q},M}(n) = L_{\mathfrak{q},M'}(n) + L_{\mathfrak{q},M''}(n) - \text{length}(M' \cap \mathfrak{q}^n M)/\mathfrak{q}^n M'.$$

But by the Artin-Rees lemma (Lemma 5.1) there is an integer m such that

$$(M' \cap \mathfrak{q}^n M) = \mathfrak{q}^{n-m}(M' \cap \mathfrak{q}^m M) \subset \mathfrak{q}^{n-m}M' \quad \text{for } n \geq m,$$

whence

$$F(n) := \text{length}(M' \cap \mathfrak{q}^n M)/\mathfrak{q}^n M' \leq L_{\mathfrak{q},M'}(n) - L_{\mathfrak{q},M'}(n-m),$$

and $F(n)$ agrees, for large n, with a polynomial of degree < the degree of the polynomial agreeing with $L_{\mathfrak{q},M'}(n)$, as required. □

We now state and prove our main theorem in a slightly more general form

Theorem 12.4. *If R is a local ring, and $\mathfrak{q}$ is a parameter ideal for the finitely generated R-module M, then*

$$\dim M = 1 + \deg P_{\mathfrak{q},M}.$$

(This number is also the degree of the polynomial agreeing with $L_{\mathfrak{q},M}(n)$ for large n).

The proof is given below. Since $P_{\mathfrak{q},M}$ depends only on the associated graded module

$$\mathrm{gr}_{\mathfrak{q}} M = M/\mathfrak{q}M \oplus \mathfrak{q}M/\mathfrak{q}^2M \oplus \ldots,$$

it does not change when we replace M by its completion (with respect to $\mathfrak{q}$ or $\mathfrak{m}$—it makes no difference). Thus, extending Corollary 10.12, we get

Corollary 12.5. $\dim M = \dim \hat{M} = \dim(\mathrm{gr}_{\mathfrak{m}} M)_P$, *where $\hat{M}$ is the completion of M with respect to $\mathfrak{m}$, and P is the ideal of elements of positive degree in* $\mathrm{gr}_{\mathfrak{m}} R$. □

Actually, the dimension of M and $\mathrm{gr}_{\mathfrak{m}} M$ coincide under quite general circumstances. See Exercise 13.8 for the case $M = R$.

The leading coefficient of $P_{\mathfrak{q},M}$ is of the form $e(\mathfrak{q}, M)/(\dim M - 1)!$ for a positive integer $e(\mathfrak{q}, M)$ called the **multiplicity** of $\mathfrak{q}$ on M; see Exercises 12.6–12.11, as well as Serre's classic book [1957] for an account of its properties. In the special case where $M = R, \mathfrak{q} = \mathfrak{m}$, and R is the localization at the "irrelevant ideal" of the homogeneous coordinate ring of a projective variety X, the multiplicity is just the degree of the projective variety X, and in general it has many properties extending those of the special case.

Proof of Theorem 12.4. We divide the proof into three steps:

Step 1: $\deg P_{\mathfrak{q},M}$ is independent of the parameter ideal $\mathfrak{q}$.

We may harmlessly replace the ring R by the ring $R/\mathrm{ann}\, M$, and thus assume that $\mathrm{ann}\, M = 0$. From the definition of parameter ideal we know that there is a number d such that $\mathfrak{m} \supset \mathfrak{q} \supset \mathfrak{m}^d$. Thus for every n we have $\mathfrak{m}^n \supset \mathfrak{q}^n \supset \mathfrak{m}^{dn}$, whence

$$L_{\mathfrak{m},M}(n) \leq L_{\mathfrak{q},M}(n) \leq L_{\mathfrak{m},M}(dn).$$

Since the outside terms are polynomials of the same degree in n, the desired assertion follows.

Step 2: $1 + \deg P_{\mathfrak{q},M} \le \dim M$.

By Proposition 10.5 there is a parameter ideal generated by $\dim M$ elements. By step 1, we may assume that it is $\mathfrak{q}$. The desired inequality now follows from Proposition 12.2.

Step 3: $1 + \deg P_{\mathfrak{q},M} \ge \dim M$.

We do induction on the dimension of M. The case $\dim M = 0$ being trivial, we may assume $\dim M > 0$.

Let P be an associated prime of M whose dimension is equal to $\dim M$. Since M contains a copy of R/P, it suffices (by the easy part of the additivity formula (Lemma 12.3)) to treat the case $M = R/P$. Since $\dim M > 0, P$ is not the maximal ideal of R and $\mathfrak{q} \not\subset P$. Any element $x \in \mathfrak{q}$ which is not in P is a nonzerodivisor on M.

It follows that $\dim M/xM < \dim M$. Using Corollary 10.9 we see that in fact $\dim M/xM = \dim M - 1$. By induction, $1 + \deg P_{\mathfrak{q},M/xM} = \dim M/xM$, and it suffices to show that $\deg P_{\mathfrak{q},M/xM} < \deg P_{\mathfrak{q},M}$. Applying the additivity formula (Lemma 12.3) to the exact sequence

$$0 \to M \to M \to M/xM \to 0,$$

we see that $P_{\mathfrak{q},M/xM}$ is a polynomial of degree $< \deg P_{\mathfrak{q},M}$, and we are done. ☐

12.2 Exercises

Exercise 12.1: Let $f \in R = k[x, y, z]_{(x,y,z)}$ be a homogeneous form of degree d, monic in x. Show that $(y, z), (y^2, z^2)$, and $(y, z)^2$ are all parameter ideals for $M = R/(f)$. Compute the corresponding Hilbert-Samuel functions.

Exercise 12.2:* Let $R = k[x, y, z, w]_{(x,y,z,w)}/I$, where I is the ideal of 2×2 minors of the matrix

$$\begin{pmatrix} x & y & z \\ y & z & w \end{pmatrix}.$$

Show that $\mathfrak{q} = (x, w)$ is a parameter ideal, and compute the Hilbert-Samuel polynomial with respect to it. Also, compute the length of $R/\mathfrak{q}$. Compare the results with Exercise 12.3, and with the Hilbert polynomial of R, computed in Exercise 1.19.

Exercise 12.3:* Show that if $\dim M = d$, and the parameter ideal $\mathfrak{q}$ is generated by d elements, then $H_{\mathfrak{q},M}(n) \le (\text{length}\, M/\mathfrak{q}M)\binom{d+n-1}{d-1}$, so that the leading coefficient of $P_{\mathfrak{q},M}(n) \le (\text{length}\, M/\mathfrak{q}M)/(d-1)!$. Further, if

the leading coefficient equals $(\text{length}\, M/\mathfrak{q}M)/(d-1)!$, then in fact

$$H_{\mathfrak{q},M}(n) = (\text{length}\, M/\mathfrak{q}M)\begin{pmatrix} d+n-1 \\ d-1 \end{pmatrix}$$

for all $n \geq 0$.

Exercise 12.4: Show that the inequality of Exercise 12.3 is not always an equality by computing the Hilbert-Samuel polynomial for the ring obtained by localizing the subring $k[s^4, s^3t, st^3, t^4] \subset k[s,t]$, at the maximal ideal (s^4, s^3t, st^3, t^4), using the parameter ideal $\mathfrak{q} = (s^4, t^4)$.

Analytic Spread and the Fiber of a Blowup

Exercise 12.5 (The growth of the numbers of generators of powers of an ideal): Let $(R, \mathfrak{m})$ be a local ring of dimension d. If $I \subset R$ is an ideal requiring r generators, then the number of generators $\mu_I(n)$ of I^n is obviously bounded by $\binom{n+r-1}{r-1}$, a polynomial of degree $r-1$ in n. By considering the blowup ring $B = R \oplus I \oplus I^2 \oplus \cdots$ and the ring of the **exceptional fiber** $S = B/\mathfrak{m}B = R/\mathfrak{m} \oplus I/\mathfrak{m}I \oplus I^2/\mathfrak{m}I^2 \oplus$, prove that in fact $\mu(n)$ agrees with a polynomial function $\nu(n)$ for large n, and that the degree of $\nu(n)$ is at most $d-1$. Show that the degree is $d-1$ if I is a parameter ideal. (The degree of $\nu(n)$ is called the **analytic spread** of I; it can be shown that it takes values between $\text{codim}(I)-1$ and $(d-1)$; see Valla [1979].)

In the geometric context, where R is the local ring of a point p on a projective variety X, and $\mathfrak{q}$ is the maximal ideal, then the projective variety associated to S is the fiber of the blowup of X at p. Exercise 12.5 says for example that if R is a regular local ring (that is, p is a nonsingular point of X) then the fiber has dimension $\dim X - 1$. In fact, the fiber is a Cartier divisor in the blowup. See Hartshorne [1977] Ch. 2.

Multiplicities

The most important invariant of a module and a parameter ideal other than the dimension is the **multiplicity**.

Exercise 12.6: Let R be a local ring, let M be a finitely generated R-module, and let $\mathfrak{q}$ be a parameter ideal for M. Let $P(n) := P_{\mathfrak{q},M}(n)$ be the Hilbert-Samuel polynomial. By Exercise 1.21, we may write $P(n)$ uniquely in the form

$$P(n) = \sum_{i=0}^{d} a_i F_i(n)$$

where $F_i(n) = \binom{n}{i}$ is the binomial coefficient regarded as a polynomial in n of degree i, the a_i are integers, and $a_d \neq 0$. The integer a_d is called the

multiplicity of $\mathfrak{q}$ on M, written $e(\mathfrak{q}, R)$. Show that $d = \dim R - 1$, that the leading term of $P(n)$ is $e(\mathfrak{q}, M)/d!$, and that $e(\mathfrak{q}, M) > 0$.

We have already encountered the multiplicity in a special case: If $R = R_0 \oplus R_1 \oplus \cdots$ is a graded ring generated in degree 1, over a field R_0, then R is the equal to the associated graded ring of the localization of R at $R_+ = R_1 \oplus R_2 \oplus \cdots$. R also corresponds to a projective variety $X \subset \mathbf{P}^r$, where $\dim_{R_0} R_1 = r + 1$. If we take $\mathfrak{q} = R_+$, then $e(\mathfrak{q}, R)$ is nothing but the degree of this projective variety.

The name *multiplicity* comes from the following case: Suppose that R is the local ring of a nonsingular variety X at a point x, P is the ideal corresponding to a subvariety Y passing through that point, and $\mathfrak{q}$ is generated by a system of parameters $z_1, \ldots, z_d$ on R/P, where $d = \dim R/P$. The statement that $\mathfrak{q}$ is a parameter ideal for R/P means that x is an isolated point of the intersection of the divisors $z_i = 0$ and Y. In this case Samuel proposed the number $e(\mathfrak{q}, R/P)$ as the correct intersection multiplicity of Y and the divisors defined by $z_i = 0$ at $x \in X$. In fact, this is enough to construct the intersection multiplicity of any collection of subvarieties, by a technique called "reduction to the diagonal"; see Exercise 13.15.

Exercise 12.7: Here are two multiplicity computations not requiring any theory. Let k be a field.

a. Let $F(x_1, \ldots, x_r)$ be a homogeneous form of degree d. Compute the degree of $R := k[x_1, \ldots, x_r]/(F)$ (the multiplicity of $\mathfrak{q} = (x_1, \ldots, x_r)$ on the local ring $k[x_1, \ldots, x_r]_{\mathfrak{q}}/(F)$). This example is surely the origin of the name "degree." Show that if k is infinite, then for a "general" choice of variables $x_1, \ldots, x_r$, the degree may be computed as length $R/(x_1, \ldots, x_{r-1})$.

b. Let R be a graded ring, finitely generated over $k = R_0$, and let $R_{(d)}$ be the Veronese subring (see Exercise 9.5). Compute the degree of $R_{(d)}$ in terms of the degree of R and the dimension of R.

We shall exhibit some of the elementary facts about the multiplicity in the next exercises. We use a strong form of Lemma 12.3, embodied in Exercise 12.10; the preceding two exercises contain useful elementary remarks that prepare for its proof.

Exercise 12.8: Suppose that M is a module over a ring R. Recall that Supp M denotes the set of primes $P \subset R$ such that $M_P \neq 0$, or equivalently such that $P \supset \operatorname{ann}(M)$. Thus two modules M and M' have the same support iff $\operatorname{rad}(\operatorname{ann}(M)) = \operatorname{rad}(\operatorname{ann}(M'))$. Suppose u is an endomorphism of M such that $u^n = 0$ for some n. Show that Supp $M =$ Supp ker $u =$ Supp coker u.

Exercise 12.9: Let $0 \to A \to B \to C \to 0$ be an exact sequence of modules over the local ring $(R, \mathfrak{m})$. Show that an ideal $\mathfrak{q} \subset R$ is a parameter ideal of B iff it is a parameter ideal of both A and of C.

Exercise 12.10: Let $0 \to A \to B \to C \to 0$ be an exact sequence of modules over the local ring $(R, \mathfrak{m})$, and suppose that $\mathfrak{q} \subset R$ is a parameter ideal of A, B, and C. Recall that $L_{A,\mathfrak{q}}(n) = \text{length } (A/\mathfrak{q}^{n+1}A)$, so that $L_{A,\mathfrak{q}}(n) = \sum_{i=0}^{n} H_{A,\mathfrak{q}}(n)$. We have seen in Lemma 12.3 that the alternating sum

$$F(n) := L_{B,\mathfrak{q}}(n) - L_{A,\mathfrak{q}}(n) + L_{C,\mathfrak{q}}(n)$$

agrees with a polynomial $Q(n)$ for large n. Following the steps indicated, prove the more precise result of Flenner-Vogel [1993].

Theorem . *With notation as above, the alternating sum $F(n)$ is a nonnegative function of n that agrees with a polynomial $Q(n)$ for large n. The two $\operatorname{gr}_{\mathfrak{q}} R$-modules $\ker(\operatorname{gr}_{\mathfrak{q}} A \to \operatorname{gr}_{\mathfrak{q}} B)$ and $\ker(\operatorname{gr}_{\mathfrak{q}} B \to \operatorname{gr}_{\mathfrak{q}} C)/\operatorname{im}(\operatorname{gr}_{\mathfrak{q}} A \to \operatorname{gr}_{\mathfrak{q}} B)$ have the same support in $\operatorname{gr}_{\mathfrak{q}} R$, and thus in particular have the same dimension δ. The degree of the polynomial Q is $\delta - 1$ (in case the modules are 0 their supports are to be interpreted as having dimension -1, and then the polynomial Q is 0).*

Steps of the Proof.

a. From the right-exact sequence

$$A/\mathfrak{q}^{n+1}A \to B/\mathfrak{q}^{n+1}B \to C/\mathfrak{q}^{n+1}C \to 0,$$

show that $L_{A,\mathfrak{q}}(n) - L_{B,\mathfrak{q}}(n) + L_{C,\mathfrak{q}}(n) = \text{length } (A \cap \mathfrak{q}^{n+1}B)/\mathfrak{q}^{n+1}A$. In particular, this shows that the alternating sum is nonnegative.

b. Consider the Rees algebra

$$\mathcal{R} := \mathcal{R}_{\mathfrak{q}}(R) = \cdots \oplus Rt^{-1} \oplus R \oplus \mathfrak{q}t \oplus \mathfrak{q}^2t^2 \oplus \cdots .$$

Set $\mathcal{R}(A) = \cdots \oplus At^{-1} \oplus A \oplus \mathfrak{q}At \oplus \mathfrak{q}^2At^2 \oplus \cdots$, and similarly for B and C. Thus, we get a (not necessarily exact) sequence of $\mathcal{R}$-modules $0 \to \mathcal{R}(A) \to \mathcal{R}(B) \to \mathcal{R}(C) \to 0$. Let $\mathcal{H}$ be the homology of this sequence in the middle. Show that $\mathcal{H}_i = [(A \cap \mathfrak{q}^i B)/\mathfrak{q}^i A]t^i$. Show that

$$\mathcal{H}/t^{-1}\mathcal{H} = \ker(\operatorname{gr}_{\mathfrak{q}} B \to \operatorname{gr}_{\mathfrak{q}} C)/\operatorname{im}(\operatorname{gr}_{\mathfrak{q}} A \to \operatorname{gr}_{\mathfrak{q}} B),$$

while the kernel of multiplication by t^{-1} on $\mathcal{H}$ is

$$\ker t^{-1} : \mathcal{H}(1) \to \mathcal{H} = \ker(\operatorname{gr}_{\mathfrak{q}} A \to \operatorname{gr}_{\mathfrak{q}} B).$$

Now show that multiplication by t^{-1} is a nilpotent endomorphism of $\mathcal{H}$, and apply Exercise 12.8. □

Exercise 12.11 (Properties of the multiplicity): The following results may all be proved as applications of Exercise 12.10.

a. Suppose we have an exact sequence $0 \to A \to B \to C \to 0$ of finitely generated R-modules on which $\mathfrak{q}$ is a parameter ideal. Show that:

i. If $\dim A = \dim B = \dim C$, then $e(\mathfrak{q}, B) = e(\mathfrak{q}, A) + e(\mathfrak{q}, C)$;

ii. If $\dim A = \dim B > \dim C$, then $e(\mathfrak{q}, B) = e(\mathfrak{q}, A)$;

iii. If $\dim A < \dim B = \dim C$, then $e(\mathfrak{q}, B) = e(\mathfrak{q}, C)$, and that one of these possibilities must be realized.

b. Suppose that for some number f, $x \in \mathfrak{q}^f - \mathfrak{q}^{f+1}$ is an element such that $x + \mathfrak{q}^{f+1} \in (\operatorname{gr}_{\mathfrak{q}} R)_f$ is not in any of the minimal primes of $\operatorname{gr}_{\mathfrak{q}} R$. Show that $\dim A/xA = \dim A - 1$, and $e(\mathfrak{q}, A/xA) = f \cdot e(\mathfrak{q}, A)$.

c. (Multiplicity as a length) Now let $d = \dim A$ and suppose that $\mathfrak{q}$ is generated by a system of parameters $x_1, \ldots, x_d$ on A. Show that if R contains an infinite field k then one may compute $e(\mathfrak{q}, A)$ as follows: Let $A_1 = A$. Having defined A_i for $i < d+1$ inductively, define $F_i \subset A_i$ to be the largest submodule of finite length. Show that some linear combination y_i of the $x_1, \ldots, x_d$, with coefficients in k, maps to an element of $\operatorname{gr}_{\mathfrak{q}} R$ that is not in any of the minimal primes of $\operatorname{gr}_{\mathfrak{q}}(A_i/F_i)$. Set $A_{i+1} = A_i/(F_i + y_i A_i)$. Show that A_{d+1} is a homomorphic image of $A/(x_1, \ldots, x_d)A$, and thus has finite length. Show that $e(\mathfrak{q}, A) = \text{length } A_{d+1}$.

The restriction in this problem to systems of parameters is not as serious as it appears: Always supposing that A contains an infinite field k, it may be shown that if $\mathfrak{q}$ is any parameter ideal for a module A of dimension d, then $\mathfrak{q}$ contains a system of parameters for A such that $e(\mathfrak{q}, A) = e((x_1, \ldots, x_d), A)$.

d. Generalize part c by showing that if $x_1, \ldots, x_d$ is a regular sequence on A, then the F_i are all zero, and $e(\mathfrak{q}, A) = \text{length } A/(x_1, \ldots, x_d)A$. (In general $e(\mathfrak{q}, A)$ can be identified as an alternating sum of lengths of modules of which this is the first term, see Serre [1957].)

e. (Linearity formula): Suppose that P is a prime ideal of A. Recall that the multiplicity of P in the primary decomposition of A was defined to be the length of the largest R_P-submodule of finite length in A_P. If P is a minimal prime of A, then A_P is itself finite length as an R_P-module. If P is not an associated prime of A, then A_P has no nonzero submodule of finite length over R_P.

Write m_P for the multiplicity of P in the primary decomposition of A. If $\mathfrak{q}$ is a parameter ideal for A, and P is an associated prime of A, show that $\mathfrak{q}$ is a parameter ideal for R/P. Show that

$$e(\mathfrak{q}, A) = \sum_{P \text{ a prime and } \dim R/P = \dim A} m(P) e(\mathfrak{q}, R/P).$$

(This formula is sometimes called the **associativity formula** for multiplicities, though the name "linearity formula" seems more descriptive.)

Hilbert Series

Exercise 12.12:* Here is a generalization of Exercise 10.12. Suppose that R is a graded ring finitely generated over an Artinian ring R_0, and let $H_R(n)$ be the length of R_n as an R_0-module. Suppose there exist homogeneous elements $x_1, \ldots, x_r$ of strictly positive degrees $d_1, \ldots, d_r$ such that $R/(x_1, \ldots, x_r)$ has finite length.

a. Show that the **Hilbert Series** $h_R(t) := \sum_{n \geq 0} H_R(n)t^n$ is a rational function of t, and that in fact $h_R(t)$ may be written as a polynomial divided by $\Pi_{i=1}^{u}(1 - t^{d_i})$; that is, it is a rational function with poles only at roots of unity.

b. Show that there is a number d (which may be taken to be the least common multiple of the degrees of the d_i) such that for each s, $H_R(dn + s)$ agrees with a polynomial in n for all $n >> 0$; that is, $H_R(n)$ is a "polynomial with periodic coefficients."

13
The Dimension of Affine Rings

In this section we shall prove Theorems A and A1, explained in Chapter 8. Theorem A1 is a form of the Noether normalization theorem, due to Nagata [1962]. It gives a kind of universal tool for the solution of many problems about affine rings. We shall illustrate this assertion by proving three other famous results: Hilbert's Nullstellensatz, Noether's theorem on the finiteness of the integral closure of an affine domain, and, in the next chapter, Grothendieck's lemma of generic freeness, with its applications to the semicontinuity of fiber dimensions.

13.1 Noether Normalization

To introduce the technique of the normalization theorem in a simple setting, we give a second proof of Corollary 10.13a.

Theorem 13.1. *If k is a field then* $\dim k[x_1, \ldots, x_r] = r$.

Proof. Let $T = k[x_1, \ldots, x_r]$. We do induction on r, the case $r = 0$ being trivial. The chain of primes

$$0 \subset (x_1) \subset \cdots \subset (x_1, \ldots, x_r)$$

has length r, so dim $T \geq r$. To prove the opposite inequality, suppose that $0 \subset P_1 \subset \cdots \subset P_m$ is any chain of distinct primes of T; we must show that $m \leq r$.

Let $f \in P_1$ be a nonzero polynomial. Lemma 13.2 will show that there are elements $x'_1, \ldots, x'_{r-1} \in T$ such that T is a finitely generated module over the k-subalgebra $S \subset T$ generated by $x'_1, \ldots, x'_{r-1}$ and f. It follows by incomparability (Corollary 4.18) that $0 \subset S \cap P_1 \subset \cdots \subset S \cap P_m$ is a chain of primes of length m. Factoring out f, we get a chain of primes of length $m-1$ in $S/(f)$. But $S/(f)$ is a homomorphic (actually isomorphic) image of a polynomial ring in $r-1$ variables, so it has dimension $\leq r-1$ by induction. Thus $m-1 \leq r-1$, and $m \leq r$ as required. □

It remains to prove the lemma that is at the heart of the whole following development. We give 3 refinements to the statement used above. The simplest and most important is statement c, and the beginner may safely ignore the other two.

Lemma 13.2. *Suppose that k is a field and that $f \in T = k[x_1, \ldots, x_r]$ is a nonconstant polynomial. There are elements $x'_1, \ldots, x'_{r-1} \in T$ such that T is a finitely generated module over the k-subalgebra generated by $x'_1, \ldots, x'_{r-1}$ and f. Further:*

a. *(Nagata [1962]) We may choose $x'_i = x_i - x_r^{e^i}$ for any sufficiently large integer e.*

b. *If f is homogeneous, then we may choose the x'_i homogeneous.*

c. *(Noether) If k is infinite then for some (in fact, for any sufficiently general) $a_i \in k$ we may choose $x'_i = x_i - a_i x_r$.*

Proof.

a. We shall show that f, written in terms of the variables $x'_1, \ldots, x'_{r-1}, x_r$, is monic in x_r of some degree d. Thus x_r satisfies a monic polynomial of degree d over the subring $S = k[x'_1, \ldots, x'_{r-1}, f]$, which one might write as $f(x_r) - f = 0$. By Proposition 4.1, T is generated as a module over S by $1, x_r, \ldots, x_r^{d-1}$.

   If we write a monomial $x_1^{a_1} \cdot \cdots \cdot x_r^{a_r}$ of f in terms of the elements $x'_1, \ldots, x'_{r-1}, x_r$, it becomes a polynomial

$$x_1'^{\,a_1} \cdot \cdots \cdot x_{r-1}'^{\,a_{r-1}} \cdot x_r^{a_r} + \cdots + x_r^{a_1 e + \cdots + a_{r-1} e^{r-1} + a_r}$$

   whose unique highest degree term is x_r^d where $d = a_1 e + \cdots + a_{r-1} e^{r-1} + a_r$. If e is greater than any of the exponents of the x_i that occur in f, then the a_i are the digits in the base-e expansion of this degree. From this we see that the degrees of the polynomials corresponding to distinct monomials of f are distinct. It follows that f, written in terms of the variables $x'_1, \ldots, x'_{r-1}, x_r$, is monic in x_r as required.

b. We shall choose $x'_1, \ldots, x'_{r-1}$ to be homogeneous elements such that $(x_1, \ldots, x_r)$ is a minimal prime over the ideal $(x'_1, \ldots, x'_{r-1}, f)$. To

do this, choose x'_i inductively using the prime avoidance lemma, Lemma 3.3: First, choose x'_1 to be any homogeneous element outside the minimal primes of $T/(f)$, and having chosen $x'_1, \dots, x'_i$, choose x'_{i+1} to be any homogeneous element outside the minimal primes of $T/(f, x'_1, \dots, x'_i)$. By the principal ideal theorem (Theorem 10.2) we know that codim $(x_1, \dots, x_r) \leq r$. By our construction no prime of codimension $< r$ can contain $(f, x'_1, \dots, x'_r)$, so that indeed $(x_1, \dots, x_r)$ is a minimal prime over the ideal $(x'_1, \dots, x'_{r-1}, f)$.

The associated primes of $(x'_1, \dots, x'_{r-1}, f)$ are all homogeneous by Proposition 3.12. It follows that $(x_1, \dots, x_r)$ is the only minimal prime of $(x'_1, \dots, x'_{r-1}, f)$, and thus $A := T/(x'_1, \dots, x'_{r-1}, f)$ is an Artinian graded ring. Thus the nth graded component A_n is 0 for large n, and A is a finite-dimensional vector space over k.

Let $S = k[y_1, \dots, y_r]$ be another polynomial ring, and let S act on T with y_i acting as x'_i for $i < r$ and as f for $i = r$. If we give S a grading by giving y_r the same degree as f and the other y_i the same degree as x'_i, then T is a graded S-module, and $T/(y_1, \dots, y_r)T = A$. By the graded Nakayama's lemma, Exercise 4.6a, T is itself finitely generated over S, as required.

c. As in part a we shall show that f, written in terms of the variables $x'_1, \dots, x'_{r-1}, x_r$, is monic in x_r. If f has degree d then it follows that T is generated by $1, x_r, \dots, x_r^{d-1}$ as a module over the subring $S = k[x'_1, \dots, x'_{r-1}, f]$.

Consider the sum f_d of all the terms of f of degree d. Writing f in terms of the $x'_i = x_i - a_i x_r$, we see that the term containing x_r^d is $f_d(a_1, \dots, a_{r-1}, 1)x_r^d$, and for sufficiently general $a_1, \dots, a_{r-1} \in k$, we shall have $f_d(a_1, \dots, a_{r-1}, 1) \neq 0$. □

Now we are ready for the Noether normalization theorem; the version stated as Theorem A1 in Chapter 8 is a special case.

Theorem 13.3 (Noether Normalization). *Let R be an affine ring of dimension d over a field k. If $I_1 \subset \cdots \subset I_m$ is a chain of ideals of R with dim $I_j = d_j$ and $d_1 > d_2 > \cdots > d_m > 0$, then R contains a polynomial ring $S = k[x_1, \dots, x_d]$ in such a way that R is a finitely generated S-module and*

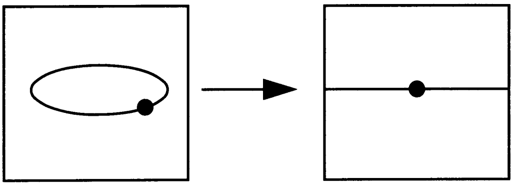
$$I_j \cap S = (x_{d_j+1}, \dots, x_d) \quad \text{for } j = 1, \dots, m.$$

If the ideals I_i are homogeneous, then the x_i may be chosen to be homogeneous. In fact, if k is infinite, and R is generated over k by $y_1, \dots, y_r$, then for $j \leq d_m$ the element x_j may be chosen to be a k-linear combination of the y_i.

The following diagram illustrates the numbering in the case $m = d$:

$$\begin{array}{rcl} k[x_1, x_2, \dots, x_d] & \subset & A \\ (x_1, x_2, \dots, x_d) & \subset & I_d \\ (x_2, \dots, x_d) & \subset & I_{d-1} \\ \dots\dots\dots\dots & & \\ (x_d) & \subset & I_1 \end{array}$$

The geometric content of Theorem 13.3, beyond that of Theorem A, is the existence, for each d-dimensional affine variety $X \subset \mathbf{A}^m$ and chain of subvarieties, of a finite map taking X to an affine space $\mathbf{A}^d$ of the same dimension, carrying the chain of subvarieties of X onto a chain of coordinate planes. The figure illustrates the case $d = 2$.



The last statement of the theorem says in particular that a suitable (and, from the proof, any sufficiently general) linear projection $\mathbf{A}^m \to \mathbf{A}^d$ induces a finite map $X \to \mathbf{A}^d$.

The algebraic statement (with no ideals I_j) is due to Noether [1926] in the case where k is infinite (though the geometric version was simply taken for granted long before this); Zariski [1943] treated the case of a finite ground field, while Nagata [1966] is responsible for the refined version given here (with a single ideal I_1).

Proof. Let $R = T/I$, where $T = k[y_1, \dots, y_r]$ is a polynomial ring. Writing I_j again for the preimage of I_j in T, and enlarging the given chain of ideals by setting $I_0 = I$, we see that it is enough to treat the case where $R = T$, a polynomial ring. In this case $r = d$ and we write

$$T = k[y_1, \dots, y_d].$$

We claim that it is enough to choose elements $x_1, \dots, x_d \in T$ such that all x_i are homogeneous in case the I_j are, and $x_1, \dots, x_{d_m}$ are k-linear combinations of the $y_1, \dots, y_d$ in case k is infinite, and

i. T is a finitely generated module over the subring S generated by $x_1, \dots, x_d$, and

ii. for each $j, I_j \cap S \supset (x_{d_j+1}, \dots, x_d)$.

To see that i and ii suffice, note first that if they are satisfied then the $x_1, \dots, x_d$ are algebraically independent over k; else by i the transcendence degree of T would be $< d$, contradicting the algebraic independence of the y_i. Thus the subalgebra of T generated by $x_1, \dots, x_d$ is isomorphic to the polynomial ring.

We next must show that the containment in ii is an equality. By Theorem 13.1 and Proposition 9.2, both ideals have dimension d_j. But the right-hand side is a prime ideal. If $I_j \cap S$ were strictly larger than $(x_{d_j+1}, \ldots, x_d)$, then $I_j \cap S$ would have smaller dimension. It follows that the two ideals are equal. This shows it is enough to find elements x_i satisfying i and ii.

To construct such x_i, we modify the elements y_i stepwise: We begin by setting $x_i' = y_i$ for $i = 1, \ldots, d$. Suppose that at a certain point we have chosen elements $x_{e+1}, \ldots, x_d$ and auxiliary elements $x_1', \ldots, x_e'$ so that

i′. T is a finitely generated module over

$$S_e := k[x_1', \ldots, x_e', x_{e+1}, \ldots, x_d].$$

ii′. for each $j, I_j \cap S_e \supset (x_h, \ldots, x_d)$, where $h = \max(d_j + 1, e + 1)$.

At the next step we shall define a new element x_e and replace $x_1', \ldots, x_{e-1}'$ with new elements (which are homogeneous if the I_j are homogeneous and are linear combinations of the $x_1', \ldots, x_e'$ if k is infinite) so that i′ and ii′ are again satisfied, with e replaced by $e - 1$. After the last step we may take $x_i = x_i'$ for $i = 1, \ldots, d_m$, completing the argument.

Suppose, then, that $x_1', \ldots, x_e', x_{e+1}, \ldots, x_d$, with $d \geq e > d_m$, have been chosen to fulfill the conditions i′ and ii′. Let j be the smallest index such that $e > d_j$. We claim —and this is the central point of the proof—that

$$I_j \cap k[x_1', \ldots, x_e'] \neq 0;$$

otherwise, since $I_j \cap S_e$ contains $(x_{e+1}, \ldots, x_d)$ by hypothesis, we would have $I_j \cap S_e = (x_{e+1}, \ldots, x_d)$. The dimension of the ideal on the left-hand side of this equality is d_j by Proposition 9.2, and the dimension of the ideal on the right-hand side is e by Theorem 13.1. Since $e > d_j$, this is a contradiction, proving our claim.

We now choose x_e to be any nonzero polynomial in $I_j \cap k[x_1', \ldots, x_e']$. If I_j and the x_i' are homogeneous then we may replace x_e' by any of its homogeneous components, and suppose that x_e is homogeneous. By Lemma 13.2 we may find elements $x_1'', \ldots, x_{e-1}'' \in k[x_1', \ldots, x_e']$ such that $k[x_1', \ldots, x_e']$ is a finitely generated module over $k[x_1'', \ldots, x_{e-1}'', x_e]$; and these elements may be taken to be homogeneous if x_e is, or of the form $x_1' - a_1 x_e', \ldots, x_{e-1}' - a_{e-1} x_e', x_e'$ in case k is infinite. The elements

$$x_1' - a_1 x_e', \ldots, x_{e-1}' - a_{e-1} x_e', x_e, x_{e+1}, \ldots, x_d$$

satisfy conditions i′ and ii′ with e replaced by $e - 1$. This concludes the proof. □

Here is the promised application to Theorem A, whose statement we recall. If $R \subset T$ are domains, we write tr. deg.$_R T$ for the transcendence degree of the quotient field of T over the quotient field of R.

Theorem A. *If R is an affine domain over a field k, then*

$$\dim R = \text{tr. deg.}_k R,$$

and this number is the length of every maximal chain of primes in R.

We postpone the proof to discuss some consequences. A first consequence is that the dimension is finite for affine rings over a field, as it is not in general for Noetherian rings.

Another is that dimension and codimension are complementary, as they ought to be, for affine rings:

Corollary 13.4. *If R is an affine domain, and $I \subset R$ is an ideal, then dim $I + \operatorname{codim} I = \dim R$.*

Proof. The dimension of R can by Theorem A be computed in terms of a maximal chain of prime ideals that includes a given minimal prime of I. □

The following corollary is a much sharper form of Theorem 10.10 for the case of affine domains.

Corollary 13.5. *If $R \subset T$ is an inclusion of affine domains over a field k, then*

$$\dim T = \dim R + \dim K(R) \otimes_R T.$$

Proof. By Theorem A we have dim $R = \text{tr. deg.}_k K(R)$, dim $K(R) \otimes_R T = \text{tr. deg}_{K(R)} K(R) \otimes_R T$, and dim $T = \text{tr. deg.}_k T$. The transcendence degree is additive, so the equality in the corollary follows. □

From a geometric point of view Corollary 13.5 says that if $Y \to X$ is a dominant morphism of varieties, then dim Y = dim X + dim (generic fiber), where we interpret the generic fiber to be a variety over $K(X)$, the field of rational functions on X. In Corollary 14.9 we shall complete this statement by showing that the dimension of the generic fiber is also the dimension of "most" fibers.

By working Theorem A a little harder we can prove a much more general version of this useful result: Namely, we may replace the condition that R is affine by the condition that R is **universally catenary**. We pause to discuss this notion.

Theorem A tells us, in particular, that in an affine domain, any two maximal ideals have the same codimension. If we go beyond affine rings, it is easy to construct localizations of affine domains where there are maximal ideals of different codimensions, and thus maximal chains of primes of different lengths, see Exercise 13.1. However, by virtue of Theorem A, any such localization still has the property that given primes $P \subset Q$, the maximal chains of primes between P and Q all have the same length. A ring R with this property is said to be **catenary,** and we say that R is **universally catenary** if every finitely generated R-algebra is catenary. Thus Theorem A implies:

Corollary 13.6. *Every field—equivalently, every affine ring—is universally catenary.*

In fact the rest of Theorem A is easy to deduce from this property. As we shall show in Corollary 18.10, virtually any ring that the geometer might meet (for example, any complete local ring) is universally catenary.

Another consequence of Theorem A allows us to compute the dimension of a graded ring by looking at only the homogeneous maximal ideals. (Intuitively, a graded ring corresponds to a cone; all components pass through the vertex, which contains the "most complicated points.")

Corollary 13.7 (Dimension of a graded ring). *Let $R = R_0 \oplus R_1 \oplus \cdots$ be a Noetherian graded ring. The dimension of R is computed as the supremum of the codimensions of the homogeneous maximal ideals. In particular, if R_0 is a field and $P_R(t)$ is the Hilbert polynomial of R, then* $\dim R = 1 + \deg P_R(t)$.

Proof. Given any maximal ideal Q of R we must show that there is a homogeneous maximal ideal Q' with $\operatorname{codim} Q' \geq \operatorname{codim} Q$. We shall prove this by induction on $\dim R_0$.

Set $Q_0 = Q \cap R_0$. It suffices to prove the assertion after localizing at the multiplicative set $R_0 - Q_0$, so we may assume that R_0 is local with maximal ideal Q_0.

Since all the minimal primes of R are homogeneous by Proposition 3.12, it suffices to prove the assertion after factoring out a minimal prime, and we may assume that R is a domain.

If $\dim R_0 = 0$ then R_0 is a field. Since R is Noetherian, it is finitely generated over R_0, so R is an affine domain. By Theorem A all the maximal ideals of R have the same codimension. For example, the codimension of Q is the same as the codimension of $Q_0 \oplus R_1 \oplus R_2 \oplus \cdots$, which is a maximal ideal that is homogeneous.

Finally, if $\dim R_0 > 0$, then Q_0 contains a nonzero element a. It suffices to prove the assertion after factoring out a. Since R_0 is a local domain, this decreases the dimension of R_0, and we are done by induction.

To deduce the second statement, note that in case R_0 is a field, the ideal $R_1 \oplus R_2 \oplus \cdots$ is the unique maximal homogeneous ideal of R. Thus it suffices to compute the dimension after localizing at this ideal. Further, the Hilbert polynomial of R is the same as the Hilbert polynomial of this local ring, so the desired result follows from Theorem 12.1. □

The conclusion of Corollary 13.7 need not hold if R is $\mathbf{Z}$-graded. For example, $k[x, x^{-1}]$ is a $\mathbf{Z}$-graded ring of dimension 1, but its only homogeneous prime ideal is 0.

The reader should compare the following statement with Corollary 13.5, of which it is a generalization and refinement, and with Theorem 10.10, which gives an inequality of the same sort. In the geometric case it is the statement that if $X \to Y$ is a map between irreducible varieties that is

dominant (that is, the image of X is dense in Y) then the dimension of X is the sum of the dimension of Y and the dimension of the generic fiber.

Theorem 13.8. *Suppose that R is a Noetherian domain, and that T is a domain containing R, finitely generated as an R-algebra. If Q is a prime ideal of T and we set $P := R \cap Q$, then*

$$\dim T_Q \leq \dim R_P + \dim K(R) \otimes_R T.$$

If R is universally catenary, and Q is maximal among those primes meeting R in P, then equality holds.

Proof. We do induction on the number of generators of T as an R-algebra. First suppose $T = R[x]/I$ for some prime ideal I in the polynomial ring $R[x]$ in one variable over R. Tensoring with R_P, we may suppose that R is local with maximal ideal P from the outset.

Note that $K(R) \otimes_R T$ is an affine domain over $K(R)$ so, by Theorem A, we have $\dim K(R) \otimes_R T = \text{tr. deg.}_R T$.

If $I = 0$ then $\text{tr. deg.}_R T = 1$. On the other hand, $\dim T_Q \leq \dim T = \dim R_P + 1$ with equality if Q is maximal among ideals containing P, by Corollary 10.13c. (This case does not use the hypothesis that R is universally catenary.)

If $I \neq 0$, then $\text{tr. deg.}_R T = 0$. Since $I \cap R = 0$ we may compute the codimension of I after localizing R at (0); that is, in the ring $K(R)[x]$. Since $K(R)[x]$ is a principal ideal domain, we get $\operatorname{codim} I = 1$. If we write Q' for the preimage of Q in $R[x]$, then

$$\begin{aligned}\dim T_Q &\leq \dim R[x]_{Q'} - \operatorname{codim} I \\ &\leq \dim R[x] - \operatorname{codim} I \\ &= \dim R + 1 - 1 \\ &= \dim R\end{aligned}$$

as required.

If now Q is maximal among primes contracting to P, then $\dim R[x]_{Q'} = \dim R + 1$, so the second inequality becomes an equality, while if R is universally catenary, then the first inequality becomes an equality, and we are done.

For the general case, suppose that $r > 1$ and that T is generated by $x_1, \ldots, x_r$ over R. Let $T' \subset T$ be the subalgebra generated by $x_1, \ldots, x_{r-1}$, and set $Q' = Q \cap T'$. By induction we have

$$\begin{aligned}\dim T'_{Q'} &\leq \dim R_P + \text{tr. deg.}_R T' \\ \dim T_Q &\leq \dim T'_{Q'} + \text{tr. deg.}_{T'} T\end{aligned}$$

and we get the inequality in the first statement of the theorem by adding these.

For the second statement of the theorem, suppose that Q is maximal among primes meeting R in P. As before, we may assume that R is local

with maximal ideal P. The hypothesis may then be restated by saying that Q generates a maximal ideal of T/PT. Applying the Nullstellensatz (Theorem 4.19) to the ring T'/PT', which is an affine ring over the field R/P, and its finitely generated algebra T/PT, we see that the preimage $Q'/PT' \subset T'/PT'$ of Q/PT is a maximal ideal, so Q' is maximal among primes of T' meeting R in P. If we assume in addition that R is universally catenary, then T' is too. Thus by induction both the above inequalities become equalities, and their sum is the equality we wanted. □

Theorem 13.8 is a form of Nagata's "altitude formula"; the usual form, given in Exercise 13.12, is easily deduced from this one. By a result of Ratliff, the conclusion of Theorem 13.8 can be used to characterize universally catenary rings; see Matsumura [1986, Section 15.3].

Proof of Theorem A. Suppose that R is an affine domain of dimension d, and let S be the polynomial subring $k[x_1, \ldots, x_d]$ as in Theorem 13.3 (with no ideals I_j specified). Clearly, tr. $\deg._k S = d$. Since R is finite over S, the quotient field of R is finite over that of S, so tr. $\deg._k R = d$ as well, proving the first statement.

Let $P_0 \subset \cdots \subset P_m \subset R$ be a chain of primes of R, so that $m \leq d$; we must show that if $m < d$, then a new prime can be inserted somewhere in the chain. Choose S as in Theorem 13.3, taking $I_j = P_j$. If $m < d$, then for some j we shall have

$$d_{j-1} := \dim P_{j-1} > d_{j-1} - 1 > \dim P_j =: d_j,$$

so the prime $Q := (x_{d_{j-1}}, \ldots, x_d)$ lies strictly between $Q_{j-1} := S \cap P_{j-1} = (x_{d_{j-1}+1}, \ldots, x_d)$ and $Q_j := S \cap P_j$. We shall finish the proof by showing that this implies the existence of a prime P strictly between P_{j-1} and P_j.

Factoring out P_{j-1} from R and Q_{j-1} from S, we may suppose that both these are zero. Note that the new S is again a polynomial ring. It suffices to find a prime P of R contained in P_j and meeting S in Q. Since S is factorial, it is normal, so the proof of Theorem A will be complete when we have proved:

Theorem 13.9. (*Going Down for integral extensions of normal rings*) *Let S be a normal domain, and let R be a domain containing S. If R is integral over S, then* ***going down*** *holds between R and S: Given primes $Q_1 \supset Q$ and a prime P_1 of R lying over Q_1, there exists a prime P of R lying over Q and contained in P_1 (Figure 13.1). This result holds even if S and R are not Noetherian.*

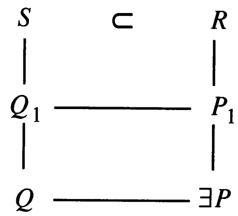


FIGURE 13.1.

The example given after Lemma 10.11 shows that we cannot drop the assumption that S is normal (even in the case of a finite extension).

Proof. Let $K(R)$ and $K(S)$ be the quotient fields of R and S. We first treat the case where $K(R)$ is a finite extension field of $K(S)$.

Let L be the normal closure of $K(R)/K(S)$ in the sense of Galois theory; that is, L is the smallest subfield of the algebraic closure $\bar{K}$ of $K(S)$ that contains $K(R)$ and is mapped into itself by any automorphism of $\bar{K}$ that fixes $K(S)$. As proved in Galois theory, the extension $L/K(S)$ is finite. Let T be the integral closure of S in L (which contains R since R is integral over S). By lying over and going up, Proposition 4.15, there are primes $P' \subset P_1'$ of T contracting to $Q \subset Q_1$ in S.

By lying over there is some prime P_1'' of T lying over P_1, and thus also over Q_1. If there is an automorphism of T fixing S and carrying P_1' to P_1'', and P'' is the image of P' under this automorphism, then the prime $P = P'' \cap R$ of R will lie over Q, as required (Figure 13.2).

The following piece of Galois theory finishes the proof of Theorem 13.9 in the case when $K(R)$ is a finite field extension of $K(S)$.

Proposition 13.10. *Let S be a normal domain with quotient field $K(S)$, let $K(S) \subset L$ be a finite normal field extension, and let T be the integral closure of S in L. If Q is a prime of S, then the primes of T lying over Q are conjugate under the Galois group of $L/K(S)$. This result holds even without the assumption that R and S are Noetherian.*

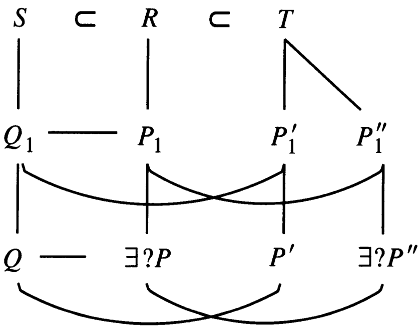


FIGURE 13.2.

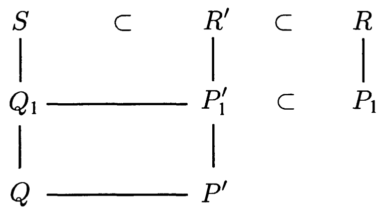


FIGURE 13.3.

As we shall see in the exercises, Proposition 13.10 can be considerably generalized, and plays a crucial role in invariant theory.

An easy special case says that if L is a purely inseparable extension of $K(S)$, then the primes of T and S are in one-to-one correspondence. Thus, from a geometric viewpoint, there are (in characteristic p only!) finite maps between nice varieties that are one-to-one and onto, but are not locally isomorphisms anywhere.

Proof. Let Q' and Q_1 be primes of T lying over Q, and let $Q_2, \ldots, Q_n$ be all the primes of T that are conjugate to Q_1. If Q' is not among the Q_i for $i \geq 1$, then by incomparability it cannot be contained in any of these Q_i. Thus by prime avoidance, Lemma 3.3, there is an $a \in Q'$ that is not in any of the Q_i. No conjugate of a can be in any of the Q_i either, so the norm $N_{L/K(S)}(a) \in S$ is outside $Q_i \cap S = Q$. On the other hand, since $a \in Q'$, the norm of a is in Q', and thus in Q; the contradiction shows that Q' must be one of the Q_i. □

Completion of the proof of 13.9. It remains to treat the case when $K(R)$ is not a finite extension of $K(S)$. Let $\bar{R}$ be the integral closure of R in $K(R)$. By lying over, Proposition 4.15, there is a prime $\bar{P}_1$ of $\bar{R}$ lying over P_1. If there exists a prime $\bar{P}$ of $\bar{R}$ contained in $\bar{P}_1$ and contracting to Q in S, then we may take $P = \bar{P} \cap R$. Thus we may assume from the outset that $R = \bar{R}$ is normal.

Now consider the set of pairs (K', P') where K' is a subfield of $K(R)$ containing $K(S)$, and P' is a prime of $R' := K' \cap R$ contained in $P_1' := P_1 \cap R'$ and contracting to Q, as in Figure 13.3. Order such pairs by setting $(K', P') \leq (K'', P'')$ if $K' \subset K''$ and $P' \subset P''$. Since an ascending union of prime ideals (in an ascending union of rings) is prime, Zorn's lemma implies that there is a maximal element (K', P'). We shall finish the proof by showing that $K' = K(R)$.

If $K' \neq K(R)$, let K'' be the result of adjoining one element of $K(R)$ to K'. Since K'' is finite over K', we may apply the case already proved to $R' \subset R'' := R \cap K''$, using the primes $P_1' \supset P'$ of R'. We see that there exists $P'' \subset R''$ contained in $P_1'' := P_1 \cap R''$ and contracting to Q in S. This contradicts the maximality of (K', P'), and thus implies that $K' = K(R)$. □

The proof of Theorem A is now complete. An important consequence is that in the case of affine domains, the local assumption may be removed from the version of the principal ideal theorem dealing with dimension given in Corollary 10.9. We have:

Corollary 13.11. *If R is an affine domain, and $f \in R$ is not a unit, then* $\dim R/(f) = \dim R - 1$.

Proof. By Theorem A we may compute both the dimensions after localizing at a maximal ideal containing f, and we may apply Corollary 10.9 to the localized rings. □

13.2 The Nullstellensatz

As a second illustration of how the Noether normalization theorem may be used, we give a very short proof of Hilbert's Nullstellensatz (already proved in somewhat stronger form in Chapter 4).

Corollary 13.12 (Hilbert's Nullstellensatz). *Let R be an affine ring over a field k, and let $P \subset R$ be a prime.*

i. If P is maximal, then R/P is a finite field extension of k.

ii. P is the intersection of maximal ideals of R.

Proof.

i. By Theorem 13.3, the 0-dimensional affine ring R/P is a finitely generated module over the polynomial ring over k in $d = 0$ variables; that is, R/P is a finite-dimensional vector space over k.

ii. If $f \in R - P$, we must find a maximal ideal of R containing P but not containing f. Factoring out P, we may assume $P = 0$. Let $S = k[x_1, \ldots, x_d] \subset R$ be a polynomial ring satisfying Theorem 13.3 with $I_1 = (f)$; since $\dim R/(f) = d - 1$ by Corollary 13.11, we have $(f) \cap S = (x_1)$. By lying over (Proposition 4.15), there exists a prime $\mathfrak{n}$ of R lying over the maximal ideal $\mathfrak{m} = (x_1 - 1, x_2, \ldots, x_d) \subset S$, and we claim that $\mathfrak{n}$ has the desired properties. Since $\mathfrak{m}$ is maximal, the dimension statement of Proposition 9.2 implies that $\mathfrak{n}$ is maximal. $\mathfrak{n}$ cannot contain f, since $\mathfrak{m}$ would then contain x_1, and with it 1. □

13.3 Finiteness of the Integral Closure

Yet another consequence of the Noether normalization theorem is the finiteness of the integral closure of an affine domain. Geometrically, the finiteness of the integral closure is the key to showing that the operation of normalization is well defined for algebraic varieties. The result remains true if we

replace affine domains by finitely generated rings (that is, finitely generated $\mathbf{Z}$-algebras); this is part of the statement that $\mathbf{Z}$ is "excellent." See Grothendieck [1965, Chapter IV 7.8, esp. 7.8.3 ii, iii, and vi].

Corollary 13.13 (Emmy Noether). *Let R be an affine domain over a field k. Set $K = K(R)$ and let L be a finite extension field of K. If T is the integral closure of R in L, then T is a finitely generated R-module; in particular, T is again an affine domain.*

It is amusing that to prove this result for the case where $L = K$, we shall reduce to the case where $L \neq K$. The idea is to use Noether normalization to replace R by a polynomial ring, which is itself normal. Once this is done, we first treat the purely inseparable case. In the separable case Proposition 13.14, a classic piece of Galois theory, does the rest. See Exercise 13.10 for the case of a complete ring.

Proof. By Noether normalization, R is a finite module over a polynomial ring $k[x_1, \ldots, x_d] \subset R$, and it suffices to prove the corollary after replacing R by this subring. Further, since a submodule of a finitely generated module is finitely generated, it suffices to prove the corollary after making a finite extension of L, and thus we may replace L by its normal closure and assume that L/K is a normal extension in the sense of Galois theory.

Let L' be the fixed field of the Galois group of L over K, so that L/L' is Galois and L'/K is purely inseparable. We shall first show that the integral closure R' of R in L' is a finitely generated R-module.

If $L' = K$ this is trivial, so we suppose that $L' \neq K$. Let p be the characteristic of L, which is necessarily nonzero since L'/K is purely inseparable. For some power q of p, the field L' is generated by qth roots of rational functions. Extending L' further by adjoining qth roots of their coefficients, we may assume that

$$L' = k'(x_1^{1/q}, \ldots, x_d^{1/q})$$

where k' is obtained from k by adjoining the qth roots of the coefficients. The integral closure of R in L is $T = k'[x_1^{1/q}, \ldots, x_d^{1/q}]$, since this ring is integrally closed, has quotient field L, and is finite over R. Since $R' \subset T$, this shows that R' is finite over R.

In Proposition 13.14 we shall see that the integral closure of the normal ring R' in the Galois extension L of L' is finitely generated over R'. Since R' is itself finitely generated over R, this completes the proof. □

Proposition 13.14. *Suppose that R is a Noetherian normal domain with quotient field K. If L is a finite separable extension of K, then the integral closure of R in L is a finitely generated R-module.*

Proof. Replacing L by its Galois closure, we may assume that L/K is Galois, with Galois group G. Let T be the integral closure of R in L, let $b_1, \ldots, b_n$ be elements of T that form a vector space basis for L/K, and let $G = \{\sigma_1, \ldots, \sigma_n\}$. Let M be the matrix over L whose (i, j)th entry is $\sigma_i b_j$,

and set $d = \det M \in L$. We shall show that $d \neq 0$ and that

$$T \subset d^{-2}(Rb_1 + \cdots + Rb_n).$$

Since $d^{-2}(Rb_1+\cdots+Rb_n)$ is a finitely generated R-module and R is Noetherian, this will show that T is a finitely generated R-module.

First, $d \neq 0$ because if the rows of M were linearly dependent then the σ_i would be linearly dependent, contradicting the linear independence of automorphisms. (Recall the argument from Galois theory, due to Artin: Suppose $\sum a_i\sigma_i = 0$ is a dependence relation with $a_i \in K$ having the smallest possible number of nonzero terms. For $b, c \in L$ we get $\sum a_i\sigma_i(b)\sigma_i(c) = \sum a_i\sigma_i(bc) = 0$, so $\sum(a_i\sigma_i(b))\sigma_i$ is also a dependence relation among the σ_i. At least two of the a_i must be nonzero; suppose $a_i \neq 0$ and $a_j \neq 0$. Choose b so that $\sigma_i(b) \neq \sigma_j(b)$. If we divide the new relation by $\sigma_i(b)$ and subtract from the old one, the terms involving σ_i cancel, but the terms involving σ_j do not. We thus get a nonzero relation having strictly fewer nonzero terms, a contradiction.)

It remains to prove that $T \subset d^{-2}(Rb_1 + \cdots + Rb_n)$. First, note that the ring T is invariant under the automorphisms of L/K since it is the integral closure of R. Thus the matrix M has entries in T. Let $b \in T$ be arbitrary, and write $b = \sum c_ib_i$ with $c_i \in K$. We must show that $d^2c_i \in R$. Let c be the column vector with ith entry c_i. Since the c_i are fixed by the automorphisms σ_i, the ith entry of the column vector Mc is $\sum_j c_j\sigma_i(b_j) = \sigma_i(\sum_j c_jb_j) = \sigma_ib \in T$. Multiplying on the left by the matrix of cofactors of M, we derive $dc_i \in T$.

Now $d = \det M \in T$. For each i, σ_id is the determinant of a matrix obtained by permuting the rows of M, so $\sigma_id = \pm d$. It follows that d^2 is invariant under G, whence $d^2 \in K$.

Since $dc_i \in T$ we now get $d^2c_i \in T \cap K$. Since R is normal, $T \cap K = R$. We have proved that $d^2c_i \in R$, or $c_i \in d^{-2}R$, as desired. □

Unfortunately, the integral closure of a Noetherian domain is generally not finite. In fact, starting in dimension 3 it need not be Noetherian. This is one of the points where the Noetherian axiomatization of what is good about affine rings is insufficient. (We have already seen that the integral closure is Noetherian in dimension 1 (Krull and Akizuki, Theorem 11.13); references and further information are given there.)

There are at least three natural responses to the problem this raises: One might choose to work, after all, only with affine rings and their relatives; in the end, they are the rings one wants to use most often. One might try to weaken the Noetherian condition and include the integral closures of all Noetherian rings among the objects of study. One attempt in this direction is the class of "Krull rings." Or one might add the finiteness of integral closure (of all factor rings that are domains) as an axiom. This approach was given great impetus by Grothendieck. He called the rings with this property **universally Japanese rings**, though the name **Nagata rings** now seems to be current. The class of Noetherian rings has other failings

as well, and Grothendieck introduced the class of **excellent rings**, defined by a list of properties that they share with affine rings, as the good class of rings to study: It contains fields and the ring of integers, and it is closed under finitely generated ring extensions, completions, and some other useful operations such as integral closures. Each of these options has its place, but the third seems to have produced the most mathematics. See Matsumura [1986] for a definition and basic properties of excellent rings.

We can use Proposition 13.14 and what we already know about complete rings and DVRs to compute the algebraic closure of the field of Laurent series in one variable over an algebraically closed field of characteristic 0. The result (in the convergent case, over **C**) and its consequences for plane curves were elaborated by Puiseux [1850], though something very similar was apparently known to Newton [1671].

Corollary 13.15. *Let k be an algebraically closed field of characteristic 0. The algebraic closure of the field $k((x))$ of Laurent series over k is the field $\cup_{n=1}^{\infty} k((x^{1/n}))$, and the integral closure of $k[[x]]$ in $k((x^{1/n}))$ is $k[[x^{1/n}]]$.*

Proof. Let L be any finite extension of $k((x))$. We shall show that the integral closure T of $k[[x]]$ in L has the form $k[[x^{1/n}]]$ for some n, and thus $L = k((x^{1/n}))$. This will prove the first statement of the corollary, and the second statement follows at once.

By Proposition 13.14, T is finite over $k[[x]]$. It follows by Corollary 7.6 that T is the direct product of complete local domains. Since T is itself a domain, T must be local. By Proposition 9.2 (with $I = 0$) T is 1-dimensional, so T is a DVR by Theorem 11.5. Write π for a generator of its maximal ideal.

Now for some n we may write $x = u\pi^n$, with u a unit of T. The residue field of T is finite over k; since k is algebraically closed, the residue field must be k. Since k is algebraically closed, the image $\bar{u}$ of u in $T/(\pi) = k$ has an nth root $\bar{v}$. Since the characteristic of k is 0, the polynomial $t^n - \bar{u}$ has a simple root at $\bar{v}$, so by Hensel's lemma, Corollary 7.4, $\bar{v}$ lifts to an nth root v of u in T. Let $\pi' = v\pi = x^{1/n}$; it is another generator of the maximal ideal of T. The map $k[[x']] \to T$ sending x' to π is an epimorphism by Theorem 7.16, and must be an isomorphism since $\dim T = 1$. □

Corollary 13.15 is often applied to plane curves by means of the following consequence due to Newton. See, for example, Walker [1978] for a discussion and a direct treatment.

Corollary 13.16. *Any polynomial equation in two variables $f(x, y) = 0$ over an algebraically closed field of characteristic 0 admits solutions of the form $y = p(x^{1/n})$ for some natural number n, where p is a Laurent series that may be taken to be a power series if f is monic in y. If in addition*

$f(0,0) = 0$, then y may be written as a power series in $x^{1/n}$ without constant term.

Proof. By Corollary 13.15 the irreducible factors of $f(x, y) = 0$ over $k((x))$ must have roots y in some $k((x^{1/n}))$. If f is monic in y, then these roots are integral and lie in $k[[x^{1/n}]]$. If $f(0,0) = 0$ then at least one of the roots y of $f(x, y)$ must reduce mod x to 0—and thus must be in the maximal ideal of $k[[x^{1/n}]]$. □

13.4 Exercises

Exercise 13.1: Let R be an affine domain, and suppose that P and Q are prime ideals of distinct codimensions, neither contained in the other. Let U be the multiplicatively closed set $R - (P \cup Q)$. Show that there are maximal chains of prime ideals in R_U of distinct lengths.

Quotients by Finite Groups

A central geometric problem is the problem of invariant theory: Given a variety X over a field k and a group G acting "algebraically" on it, find a "good" quotient X/G. As we discussed in Chapter 1, the ring of invariant functions $A(X)^G = \{f \in A(X) | f(p) = f(g(p)) \text{ for all } g \in G\}$, when it is finitely generated over k, is the best possibility for the affine coordinate ring of the quotient variety X/G. Unfortunately, in general, $A(X)^G$ is not a finitely generated algebra; and even when it is finitely generated the points of the variety corresponding to $A(X)^G$ may not be in one-to-one correspondence with the orbits of the action of G.

Neither of these bad phenomena arises in the case of finite groups, as we shall see in the next 3 problems.

Exercise 13.2:* Let G be a finite group acting on a domain T, and let R be the ring of invariants, $R = T^G$. Show that every element of T satisfies an integral equation over R—in fact, each element $b \in T$ is integral over the subring generated by the elementary symmetric functions in the conjugates σb for $\sigma \in G$.

Exercise 13.3:* Use the last remark of Exercise 13.2 to prove the following celebrated theorem of Emmy Noether [1926]:

Theorem 13.17. *If T is an affine ring over a field (of any characteristic), then the ring of invariants T^G is again an affine ring—that is, there are finitely many invariants in terms of which all others can be expressed as polynomials.*

A good deal is known about the properties of the ring of invariants in case the characteristic of the field does not divide the order of the group;

see, for example, Stanley [1979]. However, the contrary case is a wide-open field of investigation (see, for example, Peskin [1983]).

Exercise 13.4: If now the group G acts on an affine variety X with $T = A(X)$, then since T^G is again an affine ring by Theorem 13.7, we can write $T^G = A(Y)$ for some affine variety Y. Prove that the points of Y are in one-to-one correspondence with the orbits of the points of X (either assuming that the base field is algebraically closed, or interpreting points in the sense of schemes) by proving the following variation of Proposition 13.10:

Proposition 13.10′. *If T is an affine ring, and G is a finite group of automorphisms of T, then the primes of T lying over a given prime of T^G are all conjugate under G.*

Can you formulate a result that includes both Propositions 13.10 and 13.10′?

Thus the significance of Noether's theorem in Exercise 13.3 is that the quotient of an affine variety by a finite group is again an affine variety.

Exercise 13.5: The analogue of Theorem 13.3 in which dimension is replaced by codimension fails in general, for example, for the ring $R = k[y_1, y_2, y_3]/(y_1y_2, y_1y_3)$ and the ideal (y_2, y_3), but is true (and follows from Theorem 13.3) when R is an affine domain.

Primes in Polynomial Rings

Exercise 13.6: Let k be a field and let $S = k[x_1, \dots, x_r]$ be the polynomial ring. Use Theorem 13.3 to make Exercise 10.2 more explicit and make Exercise 4.27 more general as follows: Let $P \subset S$ be a prime ideal of dimension d. Set $c = d - r = \operatorname{codim} P$. Show that with respect to a suitable choice of variables $y_1, \dots, y_r$ for S, there are polynomials of the form $f_1(y_1, \dots, y_{d+1}), f_2(y_1, \dots, y_{d+2}), \dots, f_c(y_1, \dots, y_r)$ and a polynomial $g \notin P$ such that

$$P = \{f \in S | gf \in (f_1(y_1, \dots, y_{d+1}), \\ f_2(y_1, \dots, y_{d+2}), \dots, f_c(y_1, \dots, y_r))\}.$$

Dimension in the Graded Case

Exercise 13.7 (Every prime in a graded ring is nearly homogeneous):* Assume that S_0 is a field and that S is a graded ring generated as an S_0-algebra by S_1. Let $Q \subset S$ be a prime ideal, and let $P \subset Q$ be the ideal generated by the set of homogeneous elements of Q. Show that P is prime, and that either $P = Q$ or codim $Q/P = 1$ in the ring S/P. This exercise is the algebraic (and more general) version of the "obvious" geometric fact that given an irreducible algebraic set X in $\mathbf{A}^r$, there is a unique smallest cone with vertex at the origin containing it, as in Figure 13.4.

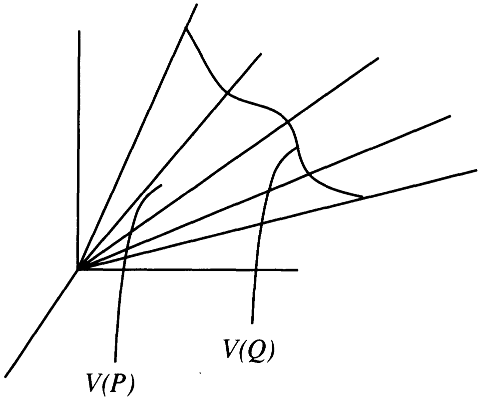


FIGURE 13.4.

This cone is obtained by drawing lines from the origin to every point of X, and thus X is of codimension 1 inside the cone.

Exercise 13.8 (Dimension of blowup and Rees algebras):* Recall that if I is an ideal of the ring R, then the blowup algebra and Rees algebras of I are the subalgebras of $R[x, x^{-1}]$ given by

$$B_I(R) := R \oplus Ix \oplus I^2x^2 \oplus \cdots$$
$$\mathcal{R}_I(R) := \cdots Rx^{-2} \oplus Rx^{-1} \oplus R \oplus Ix \oplus I^2x^2 \oplus \cdots,$$

and the associated graded algebra (called by geometers the ring of the **normal cone**) is

$$\mathrm{gr}_I(R) := R/I \oplus I/I^2 \oplus \cdots .$$

Always supposing that R is Noetherian, show:

1) a. The minimal primes of $\mathcal{R}(R, I)$ are the ideals of the form $PR[x, x^{-1}] \cap \mathcal{R}(R, I)$, where P is a minimal prime of R.

   b. $\dim \mathcal{R}(R, I) = 1 + \dim R$.

2) a. The minimal primes of $S_I(R)$ are the ideals of the form $PR[x] \cap S_I(R)$, where P is a minimal prime of R.

   b. $\dim S_I(R)$ is the maximum of the numbers $\dim R/P$, where P ranges over minimal primes containing I, and $1 + \dim R/Q$, where Q ranges over minimal primes not containing I.

3) $\dim \mathrm{gr}_I(R) = \max\{\dim R_P | P$ is a maximal ideal of R containing $I\}$.

Noether Normalization in the Complete Case

Exercise 13.9: The Noether normalization theorem is much more trivial (though no less useful) in the complete case because there Nakayama's

lemma serves as a finiteness criterion (Exercises 7.2 and 7.4). Prove a version of Theorem 13.3 in which R is replaced by a complete Noetherian local ring containing a field, and S is replaced by the power series ring over the residue class field of R.

Exercise 13.10: Prove the analogue of Corollary 13.13 for complete rings containing a field: If R is a complete local Noetherian ring containing a field, then the integral closure of R is a finitely generated R-module. (The same is true for all complete local Noetherian rings; see Grothendieck [1965].)

Exercise 13.11:* Prove that if $\varphi : X \to Y$ is a map of affine varieties with Y normal (that is, $A(Y)$ is normal) and $A(X)$ integral over $A(Y)$, then φ is open in the sense that the image of an open set is open.

Exercise 13.12 (Nagata's Altitude Formula):* Theorem 13.8 is sometimes expressed in the following form: Suppose that R is a Noetherian domain and that T is a domain containing R, finitely generated as an R-algebra. If $Q \subset T$ is a prime and $P = Q \cap R$, then $\dim T_Q \leq \dim R_P + \text{ tr. deg.}_R T - tr.deg_{R/P}(T/Q)$.

Products and Reduction to the Diagonal

Exercise 13.13 (Affine products): Suppose that $X, Y \subset \mathbf{A}^r_k$ are algebraic sets over a field k. Consider the product $Z := X \times Y \subset \mathbf{A}^{2r} = \mathbf{A}^r \times \mathbf{A}^r$.

a. Show that the ideal of Z in the big polynomial ring corresponding to $\mathbf{A}^{2r}$ is the sum of the ideals of X and Y, each written in its own set of variables. Show that the affine ring of Z may be written as the tensor product over k of the affine rings of X and Y.

b. Show that if R and T are affine rings over a field k, then

$$\dim R \otimes_k T = \dim R + \dim T.$$

Exercise 13.14 (Projective products): Given graded algebras R and T finitely generated over a field $k = R_0 = T_0$, we may consider their tensor product $R \otimes_k T$ as a graded ring with degree-d component $\sum_{i+j=d} R_i \otimes T_j$, but this is not the ring associated to the product of the projective algebraic sets corresponding to R and T. In fact:

a. Show that if $X, Y \subset \mathbf{P}^r$ are projective algebraic sets, with homogeneous coordinate rings R and T, then $R \otimes T$ is the homogeneous coordinate ring of the **join** $J(X, Y) \subset \mathbf{P}^{2r+1}$, defined as follows: Take two disjoint r-dimensional subspaces $\mathbf{P}_1$ and $\mathbf{P}_2$ of $\mathbf{P}^{2r+1}$, and regard X as embedded in $\mathbf{P}_1, Y$ as embedded in $\mathbf{P}_2$. The join $J(X, Y)$ is the union of all the lines in $\mathbf{P}^{2r+1}$ joining points of X to points of Y.

b. There is an embedding of $\mathbf{P}^r \times \mathbf{P}^r$ (as a set) into $\mathbf{P}^n$, with $n = r^2 + 2r = (r+1)(r+1) - 1$ called the **Segre embedding**, defined in

homogeneous coordinates by:

$$((a_0, \ldots, a_r), (b_0, \ldots, b_r)) \mapsto (a_0b_0, a_0b_1, \ldots, a_0b_r, a_1b_0, \ldots, a_rb_r).$$

Show that the homogeneous coordinate ring of the image is the subring of

$$k[x_0, \ldots, x_r] \otimes k[y_0, \ldots, y_r] \cong k[x_0, \ldots, x_r, y_0, \ldots, y_r]$$

generated by the bilinear forms $x_i \otimes y_j \leftrightarrow x_iy_j$. If we write $R := k[x_0, \ldots, x_r]$, and $T := k[y_0, \ldots, y_r]$, with the natural grading, the homogeneous coordinate ring of the image may also be described as the graded ring $\sum_i R_i \otimes T_i \subset R \otimes T$. (In general, the **Segre product** of two graded k-algebras $R = \sum R_i$ and $T = \sum T_i$ is the algebra $\sum R_i \otimes T_i \subset R \otimes T$.)

c. Again let $R = k[x_0, \ldots, x_r]$ and $T = k[y_0, \ldots, y_r]$. Let $S = k[\{z_{ij}\}]$ be the polynomial ring with variables z_{ij} for $0 \le i, j \le r$. Show that the kernel of the map $S \to R \otimes_k T$ sending z_{ij} to x_iy_j is the ideal of 2×2 minors of the "generic" matrix, with i, j entry z_{ij}. Show that $\dim \sum_i R_i \otimes T_i = 2r + 1$.

d. More generally, show that if $X, Y \subset \mathbf{P}^r$ are projective algebraic sets, with homogeneous coordinate rings R and T, then the image of $X \times Y \subset \mathbf{P}^r \times \mathbf{P}^r \subset \mathbf{P}^n$ has homogeneous coordinate ring $S := \sum_i R_i \otimes_k T_i \subset R \otimes_k T$. Show that $\dim S = \dim R + \dim T - 1$. Compute the multiplicity of the maximal homogeneous ideal of S (the **degree** of the algebraic set $X \times Y$ in $\mathbf{P}^n$) in terms of the corresponding data for R and T.

Exercise 13.15 (Reduction to the diagonal):

a. (Affine case) Suppose that $X, Y \subset \mathbf{A}^r$ are algebraic sets. Let $\Delta \subset \mathbf{A}^{2r}$ be the **diagonal**, that is $\Delta = \{(x, y) \in \mathbf{A}^r \times \mathbf{A}^r | x = y\}$. Show that $X \cap Y = X \times Y \cap \Delta$. Show that the ideal of Δ is generated by r linear forms.

If $\dim X + \dim Y = r$ and X intersects Y in an isolated point p, then we say that the intersection is **proper** at p. In this case the generators for Δ form a system of parameters in the local ring of $X \times Y$ at (p, p) and the multiplicity of this system of parameters is defined to be the **intersection multiplicity** of X and Y at p. This gives a way of reducing the definitions of multiplicities which we defined in Exercise 12.6, to the case of systems of parameters. This idea of reduction to the diagonal was extended from the affine case to arbitrary local rings by Serre in [1957], who used the extension to establish a homological formula defining the multiplicity directly. See Exercise A3.19 for the statement.

b. (Projective case) Consider the diagonal $\Delta \subset \mathbf{P}^r \times \mathbf{P}^r$. In terms of coordinates $x_0, \ldots, x_r$ and $y_0, \ldots, y_r$ on the two factors, show that the diagonal is defined by the 2×2 minors of the $2 \times (r+1)$ matrix with first row $x_0, \ldots, x_r$ and second row $y_0, \ldots, y_r$. In particular, its ideal is not generated by r elements, though it is of codimension r. However, show that in the local ring of $\mathbf{P}^r \times \mathbf{P}^r$ at a point of Δ this ideal of 2×2 minors is generated by r elements. For this reason the idea of "reduction to the diagonal" can still be used locally. One can also replace it by a "reduction to the join." With definitions as in Exercise 13.14a, let Δ' be the subset of $\mathbf{P}^{2r+1}$ defined by the $r+1$ linear equations $x_i = y_i$. Show that $X \cap Y = \Delta' \cap J(X, Y)$. Suppose $\dim X + \dim Y = r$ and that p is an isolated point of the intersection of X and Y. Let q be the point in $\Delta' \cap J(X, Y)$ corresponding to p. Show that the equations for Δ form a system of parameters in the local ring of $J(X, Y)$ at q. The intersection multiplicity of X and Y at p may be defined to be the multiplicity of this system of parameters. See Vogel [1984] for an interesting extension of this idea.

Equational Characterization of Systems of Parameters

Exercise 13.16: Here is a sense in which systems of parameters are like systems of indeterminates. Prove that if $(R, \mathfrak{m})$ is a local ring of dimension r and if $y_1, \ldots, y_r \in \mathfrak{m}$ is a system of parameters, then for all homogeneous polynomials $F(Y_1, \ldots, Y_r)$ with coefficients in R such that $F(y_1, \ldots, y_r) = 0$, all the coefficients of F are in $\mathfrak{m}$ by following the steps below. (The condition is not sufficient: In $k[[x, y]]$, the elements x^2, xy satisfy the condition but are not a system of parameters. For a related result see Exercise 17.16.)

a. Let $I = (y_1, \ldots, y_r)$. Show that the polynomials F as above generate the kernel of the homomorphism of rings

$$R[Y_1, \ldots, Y_r] \to B_I(R) \qquad Y_i \mapsto xy_i \in B_I(R) = R[xI] \subset R[x].$$

b. Deduce that all such F have coefficients in $\mathfrak{m}$ iff the induced epimorphism $R/\mathfrak{m}[Y_1, \ldots, Y_r] \to S_I(R)/\mathfrak{m}S_I(R)$ is an isomorphism iff dim $B_I(R)/\mathfrak{m}B_I(R) = r$.

c. Note that $B_I(R)/\mathfrak{m}B_I(R) = gr_I(R)/\mathfrak{m}gr_I(R)$. If $\mathfrak{m}$ is nilpotent mod I, then $\mathfrak{m}\,\mathrm{gr}_I(R)$ is nilpotent in $\mathrm{gr}_I(R)$. Now use part 3 of Exercise 13.8.

14

Elimination Theory, Generic Freeness, and the Dimension of Fibers

In this chapter we shall study the following question: Given a homomorphism $\varphi : R \to S$ of Noetherian rings such that S is a finitely generated R-algebra, how do the "fibers" $S \otimes_P K(R/P)$ vary as we vary the prime P of R? If S is flat over R, then as we have seen, there is some sense in which the fibers vary continuously. The main result below, Grothendieck's generic freeness lemma, a consequence of the Noether normalization theorem, implies that if $R \subset S$ are domains, then flatness always holds over a nonempty open set of R, so that "most" fibers share common properties.

Of course, we also wish to know as much as possible about how the fibers vary when they do vary. Perhaps the simplest question one could ask about the variation of the fibers is: What can be said about the set of fibers that are nonempty? Put differently, what sort of set is the image? More generally, what can be said about the set of fibers that have at least a given dimension, or exactly a given dimension? These questions are the beginning of **elimination theory**. Using the generic freeness lemma we shall prove a rather general form of the main theorem of elimination theory.

14.1 Elimination Theory

The following classical result is enormously useful.

Theorem 14.1 (Main Theorem of Elimination Theory). *If X is any variety over an algebraically closed field k, and Y is a Zariski closed subset of $X \times \mathbf{P}^n$, then the image of Y under projection to X is closed.*

We postpone the proof until we have formulated a stronger and more general version, Theorem 14.8. The outline of a quick, direct proof of Theorem 14.1 may be found in Exercise 14.1.

If $X = \mathbf{P}^m$ then Y may be defined by a collection of polynomials $f_i(x_1, \ldots, x_n, y_0, \ldots, y_m)$, homogeneous in the x_i and y_j separately. The equations for the image variety X can be found by "eliminating" the $m+1$ variables y_j from these equations, much as one eliminates a variable to solve a pair of linear equations in two unknowns. This is the origin of the name "elimination theory". Of course, in the nonlinear case, with many complicated equations, elimination may be a very difficult problem. We shall give a classical method below, and in the chapter on Gröbner bases (Chapter 15) we shall explain a technique currently used for computational solution of this problem.

As a simple example, consider the morphism $\varphi : \mathbf{P}^1 \to \mathbf{P}^2$ given by sending a point with homogeneous coordinates (s, t) to the point with homogeneous coordinates $(s^3, s^2t + st^2, t^3)$. How should we describe the image, the set of triples that can be expressed in the form $(s^3, s^2t + st^2, t^3)$? If we take the homogeneous coordinates in $\mathbf{P}^2$ to be x, y, z, the image of φ is clearly contained in the algebraic set defined by the ideal I of all the homogeneous forms $F(x, y, z)$ that vanish identically when s^3, s^2t+st^2, t^3 are substituted for x, y, z, and this is the smallest algebraic set that contains the image. In this case, I is a principal ideal generated by a single form F of degree 3. It may be described as the kernal of the map $k[x, y, z] \to k[s, t]$ sending x, y, z to s^3, s^2t+st^2, t^3, and may be obtained by "eliminating" s and t from the forms $x - s^3$, $y - (s^2t + st^2)$, $z - t^3$; essentially, that is, by taking the intersection of the ideal $(x - s^3, y - (s^2t + st^2), z - t^3)$ with the polynomial ring $k[x, y, z]$. It is not completely trivial to compute F, even in this very easy case. The answer is $F = y^3 - x^2z - 3xyz - xz^2$. Was it even obvious to the reader that the answer would be of degree 3?

In fact, the image of φ is the image under the projection $\mathbf{P}^2 \times \mathbf{P}^1 \to \mathbf{P}^2$ of the variety Y defined by the bihomogeneous ideal $(x - s^3, y - (s^2t + st^2), z - t^3)$. Thus Theorem 14.1 asserts that the image of φ is an algebraic set, and must be defined by F.

Projections from one factor of a product seem rather special among morphisms, but in fact any morphism can be put into this form. As a consequence, we have:

Corollary 14.2. *The image of a projective variety under a morphism is closed; more precisely, if Y is a projective variety over a field k and π : $Y \to X$ is a k-morphism to a projective variety X, then $\pi(Y)$ is a closed subset of X in the Zariski topology.*

Proof. Let $Y \subset \mathbf{P}^n$ be a projective variety, and let $\varphi : Y \to X$ be a morphism. Let $Z = \{(x, y) \in X \times \mathbf{P}^n | y \in Y,\ x = \varphi(y)\}$ be the graph of φ. Since the image of Y under φ is the image of Z under the projection to X, it suffices to prove that Z is closed in $X \times \mathbf{P}^n$, or equivalently in $X \times Y$.

Now X, being projective, can be embedded as a closed subset of a projective space $\mathbf{P}^m$. The product $X \times X$ is a closed set in the product $\mathbf{P}^m \times \mathbf{P}^m$; it is defined by the equations for X, repeated in each of the two sets of variables. In terms of coordinates $x_i \otimes x_j$ on $\mathbf{P}^m \times \mathbf{P}^m$, the diagonal $\Delta_{\mathbf{P}^m}$ is defined by the equations $x_i \otimes x_j - x_j \otimes x_i = 0$, so $\Delta_{\mathbf{P}^m}$ is a closed subset. The diagonal $\Delta_X \subset X \times X$ is the intersection of the diagonal $\Delta_{\mathbf{P}^m}$ of $\mathbf{P}^m \times \mathbf{P}^m$ with $X \times X$. Thus $\Delta_X \subset X \times X$ is also closed. The set Z is the preimage of Δ_X in $X \times Y$ under the morphism $1 \times \varphi$, so Z is closed as well. □

Theorem 14.1 and Corollary 14.2 are the algebraic analogues of the statement that projective varieties over $\mathbf{C}$ are compact and Hausdorff in the classical topology. The key fact in point set topology is that the image of a compact set under a continuous mapping to a Hausdorff space has closed image. Every variety is compact in the Zariski topology, but no variety other than a point is Hausdorff in that topology, so one cannot use the ideas from point set topology directly. Chevalley [1958] and Grothendieck isolated and studied the algebraic property that projective space has under the names **proper** (a kind of relative compactness) and **separated** (a relative form of the Hausdorff condition). See Eisenbud-Harris [1992] Chapter III A for an introduction to these notions, and Hartshorne [1977] Chapter II for more technical information.

Here is an illustration of the usefulness of Theorem 14.1. For others see Exercises 14.2 and 14.3.

Corollary 14.3. *Let f be a polynomial of degree d in n variables over an algebraically closed field. The condition that f can be factored nontrivially is equivalent to the vanishing of certain polynomials in the coefficients of f.*

Proof. The polymials of degree d form a vector space; let V_d be the projective space of lines in this vector space. Since a polynomial is irreducible iff a nonzero scalar multiple of it is irreducible, it makes sense to speak of the irreducibility of an element of V_d. If $d = e + f$, then the multiplication map induces a morphism $V_e \times V_f \to V_d$ whose image is obviously the set of polynomials with factors of degrees e and f. Since $V_e \times V_f$ is a projective variety, its image is closed, and the set of all reducible polynomials is a finite union of such images. □

Exactly parallel arguments show that the sets of polynomials that are perfect dth powers or possess dth powers as factors are likewise closed.

Theorem 14.1 was proved in a direct, constructive way early by Newton in special cases, and for $n = 1$ and 2 in general by Euler and Bezout around 1764. The formula we present below, equivalent to those of Euler and Bezout, was found by Sylvester in 1840. The generalization to all n occupied the attention of Cayley and a number of other nineteenth-century mathematicians. The results produced in this period involve rather complex computations and give fairly good procedures for performing elimination:

that is, for constructing, in case X is affine, the generators of an ideal of $A(X)$ defining the image of Y. In the early part of this century the tide turned, and people became less concerned with the constructive aspect of commutative ring theory, and more interested in the "general theory." Proofs like the one in Exercise 14.1 came into vogue.

This replacement of complex but constructive arguments by simple nonconstructive ones goes under the name of "elimination of elimination theory" (Weil, in his influential book [1946, p. 31], says, "The device that follows ..., it may be hoped, finally eliminates from algebraic geometry the last traces of Elimination-Theory ... .")[1] It has been pointed out, notably by Abhyankar, that one loses interesting information if one ignores the constructive methods. He suggested in a famous poem that one should rather

Eliminate, eliminate, eliminate,

Eliminate the eliminators of elimination theory.

Whatever the merits of this argument, the advent of computers has renewed interest in finding efficient algorithms for performing elimination. The most effective current algorithms do not follow the older methods, but are based on the theory of Gröbner bases, explained in Chapter 15.

Before proceeding to a nonconstructive general proof of Theorem 14.1, it is worth understanding a constructive proof in the simplest case. Consider an affine variety X and an algebraic subset Y of $X \times \mathbf{P}^1$. Let x_0, x_1 be coordinates on $\mathbf{P}^1$, so that Y is defined by homogeneous polynomials in x_0, x_1 with coefficients in $A(X)$. We shall suppose that Y is actually defined by exactly two such polynomials, f and g, of degrees d and e in x_0, x_1 (it is in any case not hard to reduce to this situation).

First suppose that $d = e = 1$, that is $f(s, x_0, x_1) = f_0(s)x_0 + f_1(s)x_1$, and $g(s, x_0, x_1) = g_0(s)x_0 + g_1(s)x_1$, where the $f_i(s)$ and $g_i(s)$ are functions in $A(X)$. We claim that the image of Y under projection to X is the closed set defined by the vanishing of the determinant of the matrix

$$M = \begin{pmatrix} f_0(s) & f_1(s) \\ g_0(s) & g_1(s) \end{pmatrix},$$

as a function of $s \in X$. Indeed, if s is in the image of a point $(s, u, v) \in Y$ with $u, v \in k$, then multiplying the first column of the above matrix by u and the second by v and adding, we get the column vector with components $(f(s, u, v), g(s, u, v)) = (0, 0)$, so the determinant vanishes. Conversely, if the determinant vanishes for some $s \in X$, then we may reverse the argument, or, more suitably for generalization, we may note that if the two rows of the matrix are linearly dependent for some value of s, then for this s the functions $f(s, x_0, x_1)$ and $g(s, x_0, x_1)$ are linearly dependent as polynomials

[1] Reprinted from Foundations of Algebraic Geometry by Andre Weil, Colloquium Publications, Vol. 29, p. 31, by permission of the American Mathematical Society.

$$M = \begin{pmatrix} f_0 & f_1 & f_2 & 0 & 0 \\ 0 & f_0 & f_1 & f_2 & 0 \\ 0 & 0 & f_0 & f_1 & f_2 \\ g_0 & g_1 & g_2 & g_3 & 0 \\ 0 & g_0 & g_1 & g_2 & g_3 \end{pmatrix}$$

FIGURE 14.1.

in x_0, x_1. Thus $f(s, x_0, x_1)$ and $g(s, x_0, x_1)$ share a common zero (u, v). It follows that $(s, u, v) \in Y$, and s is in the image of Y.

Sylvester generalized this argument to the case where d and e are arbitrary: We take M to be the $(d + e) \times (d + e)$ matrix whose first e rows contain the coefficients of the polynomials

$$x_0^{e-1-i} \cdot x_1^i \cdot f \qquad i = 0, \ldots, e - 1$$

and whose last d rows contain the coefficients of the polynomials

$$x_0^{d-1-i} \cdot x_1^i \cdot g \qquad i = 0, \ldots, d - 1.$$

For example, if

$$f = f_0 x_0^2 + f_1 x_0 x_1 + f_2 x_1^2$$
$$g = g_0 x_0^3 + g_1 x_0^2 x_1 + g_2 x_0 x_1^2 + g_3 x_1^3,$$

then M is as in Figure 14.1, where as before $f_i = f_i(s)$ is a function on X, and similarly for the g_i.

We claim that the vanishing of $\det(M)$, as a function on X, defines the image of Y in X. (Sylvester simply asserted this fact; a proof was later supplied by Cauchy.) As before, a point (s, u, v) of Y over $s \in X$ gives rise to a linear relation on the columns of M, showing that $\det(M)(s) = 0$; while if $\det(M)(s) = 0$ for some particular $s \in X$, then the linear dependence of the rows of M shows that $f(s, u, v)$ and $g(s, u, v)$ satisfy a relation of the form

$$a(x_0, x_1) f(s, x_0, x_1) = b(x_0, x_1) g(s, x_0, x_1),$$

where the degree of a is less than e and the degree of b is less than d. Thus some root (u, v) of $g(s, x_0, x_1)$ is not a root of a (or at least occurs with lower multiplicity), and thus is a root of $f(s, x_0, x_1)$; it follows that (s, u, v) is a point of Y over s, establishing the claim.

We now leave this tour of the beginning of elimination theory, and turn to the technical business of the chapter.

14.2 Generic Freeness

The following result is often referred to as the "generic flatness lemma" though its conclusion is that a certain module is free, a condition stronger than flatness. The extra strength is required by many applications.

Theorem 14.4 (Grothendieck's Generic Freeness Lemma). *Suppose that R is a Noetherian domain and S is a finitely generated R-algebra. If M is a finitely generated S-module, then there exists an element $0 \neq a \in R$ such that $M[a^{-1}]$ is a free $R[a^{-1}]$-module. If in addition, $S = S_0 \oplus S_1 \oplus \cdots$ is positively graded, with R acting in degree 0, and if M is a graded S-module, then a may be chosen so that each graded component of $M[a^{-1}]$ is free over R.*

The proof is a classic example of a technique Grothendieck called *dévissage.* (English: "unscrewing." After one application of the recursive step of the argument, we are back to the same spot but one dimension lower.)

Proof. Let $K = K(R)$ be the quotient field of R. We do induction on $d := \dim K \otimes_R S$, starting with the case where $K \otimes_R S = 0$, which we may think of as having dimension $d = -1$. In this case $1 \in S$ is annihilated by some nonzero element $a \in R$, and it follows that a annihilates M, whence $M[a^{-1}] = 0$, and the theorem is trivially satisfied.

Since $K \otimes_R S$ is a finitely generated algebra over the field K we may apply the Noether Normalization theorem (Theorem 13.3) in the case with no ideals. Thus there exist algebraically independent elements $x_1, \ldots, x_d$ of $K \otimes_R S$ such that $K \otimes S$ is a finitely generated module over $K[x_1, \ldots, x_d]$. If S is graded as above, then as Theorem 13.3 allows, we choose the x_i to be homogeneous. Multiplying each x_i by a suitable element of R, we may assume that each $x_i \in S$. Let $b_1, \ldots, b_t$ generate S as an R-algebra. Each b_i satisfies an integral equation over $K[x_1, \ldots, x_d]$. Clearing denominators, we may write this as a polynomial equation with coefficients in R and leading coefficient c_i, say. Let $a = \Pi c_i$. It follows that $S[a^{-1}]$ is integral, and thus is a finitely generated module, over $S' := R[a^{-1}][x_1, \ldots, x_d]$. Of course $M' := M[a^{-1}]$ is then a finitely generated module over S' as well.

By Proposition 3.7 there exists a finite filtration of M' by S'-submodules $M' = M'_1 \supset \ldots \supset M'_{s+1} = 0$ with successive quotients $M'_i/M'_{i+1} \cong S'/Q_i$, where each Q_i is a prime of S'. If $Q_i \neq 0$, then $\dim K \otimes_R S'/Q_i < d$, so by induction there is an element $a_i \in R$ such that $S'/Q_i[a_i^{-1}]$ is free over $R[a_i^{-1}]$. If $Q_i = 0$, then $S'/Q_i = S'$ is a free $R[a^{-1}]$-module (the free basis is the set of monomials in the x_i), and we set $a_i = a$. Over the ring $R[(a_1 a_2 \cdots a_s)^{-1}]$ the module $M[(a_1 a_2 \cdots a_s)^{-1}]$ has a finite filtration by free R-modules and is thus free as required. If S and M are graded as above, then by Proposition 3.12 the Q_i may be taken homogeneous. In this case the homogeneous components of $S'/Q_i[a^{-1}]$ are free, over R, and the assertion for the graded case follows. □

14.3 The Dimension of Fibers

Recall that if $\varphi : R \to S$ is a ring homomorphism then the **fiber** of φ at a prime P of R is $K(R/P) \otimes_R S$. If R and S are affine coordinate

rings of algebraic sets X and Y, respectively, φ comes from a morphism $F : Y \to X$, and if P is the maximal ideal of R representing a point p of X, then $K(R/P) \otimes_R S = S/PS$ is the affine coordinate ring of the (scheme-theoretic) fiber $F^{-1}(p)$, whence the terminology.

Using generic freeness we can go much more deeply into the study of the fibers of a homomorphism of rings. Here we shall prove some results that bear on dimension theory. Perhaps the most important further result shows that under good circumstances most fibers are "smooth" in an appropriate sense; it will be treated in Corollary 16.23, after we have the theory of differentials.

To get an idea of what to expect, consider the map $F : Y = \mathbf{A}^2 \to X = \mathbf{A}^3$ given by $(u, v) \mapsto (u, uv, 0)$ (see Figure 14.2). If we take coordinates x, y, z on the target $\mathbf{A}^3$, then this corresponds to the map of rings $R = k[x, y, z] \to k[u, v] = S$ sending $x \mapsto u, y \mapsto uv, z \mapsto 0$. We may describe the image as the (x, y)-plane, minus the y-axis, with the origin put back in.

We may write the image as the union of the open subset defined by $x \neq 0$ of the closed set defined by $z = 0$, and the closed set defined by $x = y = z = 0$. If we define a **locally closed set** to be an open subset of a closed subset (equivalently: the intersection of an open subset and a closed subset) and a **constructible set** to be a finite union of locally closed sets, then we see that the image in Figure 14.2 constructible. Chevalley's theorem, proved below, says that this is true in general, and more generally that the image of a constructible set is again constructible. The subsets X_d of X where the fibers have dimension $\geq d$ are similarly constructible, as is the difference $X_d - X_{d+1}$ where the fiber has dimension $= d$.

The situation becomes much simpler when viewed from the perspective of the source variety $Y = \mathbf{A}^2$ instead of the target $X = \mathbf{A}^3$. First, *every* point of Y has an image in X, so the set of points $q \in Y$ such that the fiber through q has dimension ≥ 0 is the whole of Y, and is thus closed.

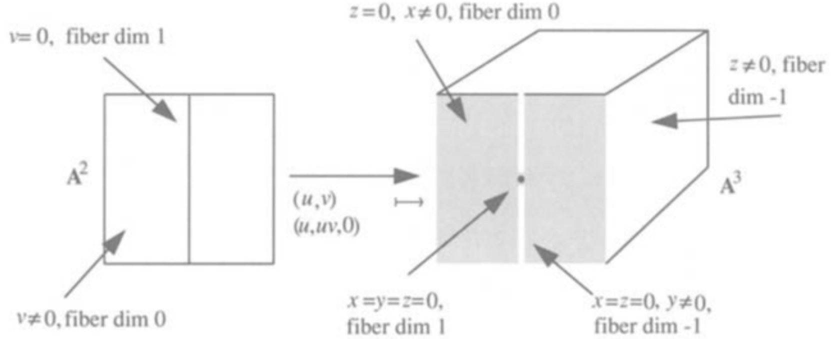


FIGURE 14.2. The image of an affine variety is constructible, not closed; fiber dimension is semicontinuous in the source, not in the target.

This trivial fact is actually typical: If we define Y_d to be the set of points $q \in Y$ such that $F^{-1}F(q)$, the fiber through q, has dimension $\geq d$, then we see that in the example each Y_d is closed in Y. Informally, one might say that as a point q_t approaches $q = q_0$, the dimension of the fiber through q_t can jump up at $t = 0$ but not down. This is expressed by saying that the dimension of fibers is **upper semicontinuous** on the source. (More formally, an invariant is **upper semicontinuous** if for each integer d, the set on which the invariant takes values $\geq d$ is closed.)

We shall prove this upper semicontinuity below. It follows from the main theorem of elimination theory that in the case of a map from a projective variety, the fiber dimension is actually semicontinuous on the target. We shall reverse this implication and prove a strong result about graded rings from which the semicontinuity theorems and the main theorem follow. First we give a special case that follows directly from the generic freeness lemma and will be useful in the proof. It uses Theorem 14.4 to sharpen Theorem 10.10 and gives the promised geometric version of Theorem 13.8.

Corollary 14.5. *Suppose that R is a Noetherian domain and that S is a finitely generated R-algebra containing R. There is an element $0 \neq a \in R$ such that for any prime ideal $P \subset R$ not containing a there are prime ideals $Q \subset S$ with $P = R \cap Q$, and for any such Q we have:*

$$\dim S_Q = \dim R_P + \dim S_Q/PS_Q.$$

Proof. By Theorem 14.4 we may choose $a \in R$ such that $S[a^{-1}]$ is free over $R[a^{-1}]$. Suppose that $P \subset R$ is a prime not containing a. The local ring R_P is a further localization of $R[a^{-1}]$, so

$$R_P \otimes_R S = R_P \otimes_{R[a^{-1}]} S[a^{-1}]$$

is free as an R_P-module. It follows that $P(R_P \otimes S) \neq R_P \otimes S$; thus it is contained in a prime ideal Q' whose intersection with S is the desired Q. The further localization $(R_P \otimes S)_{Q'} = S_Q$ is flat over $R_P \otimes S$, and thus also flat over R_P. We may now apply Theorem 10.10 to get the dimension equality. □

Specializing still further, and using Theorem A, we obtain the form that is most often used.

Corollary 14.6. *Let $R \subset S$ be an inclusion of affine domains over a field k. Set $d =$ tr. deg.$_R S$. There is an element $a \in R$ such that for each maximal ideal $P \subset R$ not containing a there exist primes Q of S containing P, and for any prime Q minimal over PS we have* $\dim S/Q = d$.

Proof. Choose $a \in R$ as in Corollary 14.5. Let P be a maximal ideal of R not containing a. Let $Q, Q_1, \ldots, Q_s$ be the minimal primes of PS. By prime avoidance we may for each i choose an element $f_i \in Q_i$ but not in Q. By the

Nullstellensatz Q is an intersection of maximal ideals, so there is a maximal ideal Q' of S that contains Q but not the element $\prod_i f_i$. By Corollaries 14.5 and 13.5 we have $\dim S_{Q'}/PS_{Q'} = \dim S_{Q'} - \dim R_P = d$. By construction, Q' does not contain any of the ideals Q_i, so $\dim S_{Q'}/PS_{Q'} = \dim S_{Q'}/Q_{Q'}$. By Theorem A dim $S_{Q'}/Q_{Q'} = \dim S/Q$. Putting these together we get $d = \dim S/Q$ as required. □

Here is the geometric version of Corollary 14.6, in the case where k is algebraically closed. Let $\varphi : Y \to X$ be a dominant morphism of affine varieties over an algebraically closed field k. There is a Zariski open subset U of X such that for each $x \in U$ the fiber $\varphi^{-1}(x)$ is nonempty, and all its components have dimension $\dim Y - \dim X = \text{tr. deg.}_{K(R)} K(S)$.

The result showing that the image of a morphism of varieties is constructible has the following algebraic expression.

Corollary 14.7 (Chevalley's Theorem). *If R is a Noetherian ring and $f : R \to S$ is a homomorphism of rings making S into a finitely generated R-algebra, then the set of primes*

$$X(f) = \{P \in \operatorname{Spec} R|\ \textit{there exists a prime } Q \textit{ of } S \textit{ with } f^{-1}(Q) = P\}$$

is constructible in Spec R.

Proof. For any ideal I of R we let $f_I : R/I \to S/IS$ be the may induced by f. By Noetherian induction we may assume that $X(f_I)$ is constructible whenever I is nonzero.

The ring S has finitely many minimal primes Q_i. We have $X(f) = \cup X(f_i)$ where $f_i : R \to S/Q_i =: S_i$ is the induced map. Thus it suffices to show that $X(f_i)$ is constructible. Since $X(f_i)$ is a subset of the spectrum of $R_i := R/f_i^{-1}(Q_i)$, it suffices to treat the case where R and S are domains and $R \subset S$.

In this case, let $0 \neq a \in R$ be as in Corollary 14.5. The open set X_1 of primes not containing a is in $X(f)$, so it suffices to treat the intersection of $X(f)$ with the complement of X_1.

The complement, $\operatorname{Spec} R - X_1$, may be identified with $\operatorname{Spec} R/(a)$. Its intersection with $X(f)$ is the set $X(f')$, where $f' : R/(a) \to S/Sa$ is the induced map. This is a constructible set by the Noetherian induction. □

We now turn to the general semicontinuity results.

Theorem 14.8 (Semicontinuity of fiber dimension). *Suppose that R is a Noetherian ring and that S is a finitely generated R-algebra. For each integer e:*

a. *There is an ideal $I_e \subset S$ such that if Q is a maximal ideal of S and $P := R \cap Q$, then*

$$\dim S_Q/PS_Q \geq e \text{ iff } Q \supset I_e.$$

b. *If $S = S_0 \oplus S_1 \oplus \cdots$ is a positively graded algebra, finitely generated over $R = S_0$, then there is an ideal J_e of R such that, for any prime ideal $P \subset R$,*

$$\dim K(R/P) \otimes S \geq e \text{ iff } P \supset J_e.$$

In geometric terms, Theorem 14.8 deals with the fibers of a morphism $Y \to X$. Part b shows that if Y is projective over X then the fiber dimension is upper semicontinuous on the target, not just on the source. Part a shows that for each e, the union of those irreducible components of fibers that have dimension $\geq e$ is a closed set in Y. Part a does *not* follow directly from part b, even in the case when Y is projective over X: Although part b says that the union of the fibers of dimension $\geq e$ is a closed subset of Y, it is not subtle enough to distinguish among the components of the fibers.

The proofs require correspondingly different tools. Both use the generic freeness lemma in the form of Corollary 14.5. But though we can deduce part b directly from the characterization of dimension by Hilbert polynomials, we shall use Theorem 13.8 for part a; we thus prove it only in the case where the ring R is universally catenary (a case that covers virtually all geometric applications—see Theorem A and Corollary 18.10). The general case may be reduced to this one, essentially by transferring the problem to finitely generated subrings. For this reduction see Grothendieck [1966, Section 13.1], where the result is proved even when R is not Noetherian, assuming that S is a finitely presented R-algebra.

Proof. The theorem is certainly true if R is a field. Working in either the graded or the ungraded case, we do "Noetherian induction" as follows: If the theorem is false for R and some R-algebra S, then there is an ideal I in R maximal among those such that the result is false for R/I and some R/I-algebra S'. Replacing R by R/I, we may assume that for any nonzero ideal $J \subset R$, the result is true for R/J and any R/J-algebra S'. In particular, we may assume that $R \subset S$—else we factor out the kernel of the map $R \to S$ and use the inductive hypothesis above.

Proof of Part a (assuming R is universally catenary). For any prime Q of S we have

$$\dim S_Q/PS_Q = \max_{Q' \text{ a minimal prime of } S} \dim S_Q/(PS_Q + Q').$$

Thus the set of Q such that the left-hand side $\geq e$ is the union, over the set of minimal primes Q', of the set of Q containing Q' and having $\dim S_Q/(PS_Q + Q') \geq e$. For this reason it suffices to prove the theorem after factoring out one of the minimal primes of S and its preimage in R, and we may assume that both R and S are domains.

Let $d = \dim K(R) \otimes S$. For every maximal $Q \subset S$ we have

$$\dim S_Q/PS_Q \geq \dim S_Q - \dim R_P = d$$

by Theorems 10.10 and 13.8. Thus we may take $I_d = 0$.

Let $0 \neq a \in R$ be chosen as in Corollary 14.5, so that if $a \notin Q$ then $\dim S_Q/PS_Q = \dim S[a^{-1}] - \dim R[a^{-1}] = d$. We see that any maximal ideal Q such that $\dim S_Q/PS_Q > \dim S - \dim R$ must contain a.

By the induction, the result is true for the R/Ra-algebra S/Sa. If $I'_e \subset S/Sa$ is the ideal corresponding to this induced map, then for $e > d$ we may take I_e to be the preimage in S of I'_e. This concludes the proof of part a.

Proof of Part b. If R were not a domain, then by induction there would be for every minimal prime P of R an ideal $J_{e,P}$ containing P such that

$$\dim K(R/P) \otimes S \geq e \text{ iff } P \supset J_{e,P} \text{ for primes } P \text{ of } S \text{ containing } PS.$$

If follows that we could take $I_e = \cap_{P \text{ a minimal prime of } R} I_{e,P}$. Thus we may suppose that R is a domain.

By Corollary 13.7 we may compute $\dim K(R/P) \otimes S$ for any prime P of R from the degree of the Hilbert polynomial that agrees with the numerical function $\dim_{K(R/P)} K(R/P) \otimes S_n$ for large n. By Nakayama's lemma, $\dim_{K(R/P)} K(R/P) \otimes S_n$ is the number of elements required to generate the R_P-module $(S_n)_P$, so that $\dim_{K(R/P)} K(R/P) \otimes S_n \geq \dim_{K(R)} K(R) \otimes S_n$. Thus for all primes $P \subset R$, we have $\dim K(R/P) \otimes S \geq \dim K(R) \otimes S$, and setting $d = \dim K(R) \otimes S$ we may take $I_d = 0$.

Let $0 \neq a \in R$ be chosen as in the second part of Theorem 14.4, so that if $a \notin Q$ then each $S_n[a^{-1}]$ is free over R. It follows that for each prime P of R that does not contain a, the R_P module $(S_n)_P$ is free of rank equal to $\dim_{K(R)} K(R) \otimes S_n$. Thus the Hilbert polynomial for $K(R/P) \otimes S$ is the same as that for $K(R) \otimes S$, so $\dim K(R/P) \otimes S = \dim K(R) \otimes S$. We see that any prime $P \subset R$ such that $\dim K(R/P) \otimes S > \dim K(R) \otimes S$ must contain a.

By induction the result is true for the R/Ra-algebra S/Sa. If $I'_e \subset S/Sa$ is the ideal corresponding to this induced map, then for $e > \dim K(R) \otimes S$ we may take I_e to be the preimage in S of I'_e. This completes the proof of part b. □

Let X be an affine variety, and let Y be a closed subset of $X \times \mathbf{P}^n$. If R is the affine coordinate ring of X, then Y corresponds to a graded R-algebra $S = S_0 \oplus S_1 \oplus \dots$ with $S_0 = R$, generated over R by $n+1$ elements of S_1. The points of $Y = \operatorname{Proj} S$ correspond to the maximal homogeneous primes of S not containing the irrelevant ideal $S_+ = S_1 \oplus S_2 \oplus \cdots$. Thus the image of Y in X corresponds to the set of maximal ideals P of R such that the fiber algebra S/PS has some homogeneous prime ideals that do not contain the "irrelevant ideal" $S_1/PS_1 \oplus S_2/PS_2 \oplus \cdots$. Since P is maximal, the irrelevant ideal is a maximal ideal of S/PS, and the condition for there to be homogeneous primes that do not contain it is $\dim S/PS \geq 1$. More generally, the dimension of the fiber of $Y \to X$ over a maximal ideal P of X is one less than the dimension of the fiber algebra S/PS. (Those who know about schemes will have no trouble interpreting and verifying the

same thing for arbitrary primes P of R.) Applying this argument in the special case where R is an affine domain over an algebraically closed field gives:

Corollary 14.9. *Let X be any variety over an algebraically closed field k, and let Y be a Zariski closed subset of $X \times \mathbf{P}^n$. For any number e, if X_e is the set of points p of X such that the fiber of Y over p has dimension $\geq e$, then X_e is closed in X.*

14.4 Exercises

Elimination Theory

Exercise 14.1 (Proof of the Main Theorem of Elimination Theory): Prove Theorem 14.1 by following these steps (if you get stuck you can find this proof in Mumford [1976]):

1. Reduce to the case $X = \mathbf{A}^m$.

2. Suppose the set Y is defined by a collection of polynomial equations

$$f_i(x_1, \ldots, x_m; y_0, \ldots, y_n) = 0$$

that are homogeneous in the second set of variables. Show that the fiber over a point $a \in \mathbf{A}^m$ is empty iff the polynomials $f_i(a; y_0, \ldots, y_n)$ generate an ideal I in $k[y_0, \ldots, y_n]$ that contains some power of the "irrelevant ideal" $(y_0, \ldots, y_n)$.

3.* Let $X_d \subset \mathbf{A}^m$ be the subset containing those points a for which I does not contain the dth power of $(y_0, \ldots, y_n)$, so that the image of Y in X is $\cap_d X_d$. Show that each X_d is closed by exhibiting defining equations.

Exercise 14.2 (Liouville's Theorem): Show that there are no nonconstant functions on a projective variety; thus the image of any map from a projective variety to $\mathbf{A}^m$ is a point. This is the algebraic analogue of a consequence of the maximum modulus principal in complex analytic geometry.

Exercise 14.3: Here is a classic method for proving that an algebraic set is irreducible. Suppose that X is an variety (=irreducible algebraic set) and $Y \subset X \times \mathbf{P}^n$ is an algebraic subset, all over an algebraically closed field. Suppose that the projection map $\pi_1 : Y \to X$ to the first factor has fibers that are all irreducible and of constant dimension. Use the Theorem on the upper-semicontinuity of fiber dimension to show that Y is irreducible.

Show that some of the hypotheses of this result are necessary by giving examples of:

a. A reducible algebraic set $Y \subset X \times \mathbf{A}^n$ where X is irreducible and such that π_1 has irreducible fibers, all of the same dimension.

b. A reducible algebraic set $Y \subset X \times \mathbf{P}^n$ where X is irreducible and such that π_1 has irreducible fibers of varying dimensions.

Exercise 14.4: Find the dual of the conic defined by the equation $x^2 + y^2 + z^2 = 0$ in $\mathbf{P}^2$.

Exercise 14.5: Suppose $\varphi : Y \to X$ is a morphism of varieties corresponding to an inclusion of affine coordinate rings $S \hookleftarrow R$, and that S is integral over R. Show that for any variety Z and any closed subset $W \subset Y \times Z$, the image of W under the morphism $\varphi \times 1 : Y \times Z \to X \times Z$ is Zariski closed in $X \times Z$. Thus such morphisms have the same property that is exhibited in Theorem 14.1 in the case of the map from a projective space to a point—this property is called **properness.**

Exercise 14.6: Suppose that R is a Noetherian domain with quotient field K, and that S is a finitely generated R-algebra. Use generic freeness to show that if M is a finitely generated S-module such that $K \otimes_R M = 0$, then there exists an element $0 \neq a \in R$ such that $R[a^{-1}] \otimes_R M = 0$. Taking $R = \mathbf{Z}$, find an example of a countably generated $\mathbf{Z}$-module S such that $\mathbf{Q} \otimes S = 0$ but $\mathbf{Z}[n^{-1}] \otimes S \neq 0$ for every $n \in \mathbf{Z}$.

Exercise 14.7 (Strong form of Chevalley's Theorem): If R is a Noetherian ring, S is a finitely generated R-algebra by a map $f : R \to S$, and Z is a constructible subset of $\operatorname{Spec} S$, show that $\{P \subset \operatorname{Spec} R | P = f^{-1}(Q) \text{ for some } Q \in Z\}$ is constructible. If $Y \to X$ is a morphism of affine varieties over an algebraically closed field, show that the image of a constructible subset of Y is a constructible set in X.

Exercise 14.8: The set of hyperplanes in $\mathbf{P}^n$ forms a projective space $\mathbf{P}^{n\vee}$, where the homogeneous coordinates of a hyperplane are taken to be the coefficients of the linear form vanishing on the hyperplane. Given an algebraic set $X \subset \mathbf{P}^n$, consider the "universal hyperplane section" of X, which is the set $Y = \{(x, H) \in \mathbf{P}^n \times \mathbf{P}^{n\vee} \mid x \in X \cap H\}$. (The name comes from the fact that if we let $\pi_2 : Y \to \mathbf{P}^{n\vee}$ be the second projection, then the fibers of π_2 are the hyperplane sections of X.) Show that Y is an algebraic subset of $\mathbf{P}^n \times \mathbf{P}^{n\vee}$. Use the first projection $\pi_1 : Y \to X$ to compute the dimension of Y. Use Exercise 14.3 to show that if X is irreducible then Y is irreducible too. Describe the ideal of Y by giving polynomials whose radical is the ideal of Y.

15
Gröbner Bases

> Man kann dieses Verfahren dazu benützen, den Restkalssenring eines nulldimensionalen Polynomideals wirklich zu berechnen ....
>
> (One can use this process to actually compute a zero-dimensional residue class ring of a polynomial ring ....)
>
> —W. Gröbner [1939]

We shall work throughout this chapter with a polynomial ring $S = k[x_1, \ldots, x_r]$ over a field k. The elements of k will be called **scalars**. All S-modules mentioned will be assumed finitely generated.

A great deal of modern commutative algebra and algebraic geometry is formulated in an essentially nonconstructive fashion. To take a simple example, Hilbert's basis theorem assures us that there exists a finite basis of the syzygies for any finite set of elements of S, but at first glance it would seem that one must investigate syzygies of all degrees to find such a basis. Nevertheless, one can find such a basis algorithmically (Hilbert's original proof was algorithmic!), and one can effectively perform a very large proportion of the other central operations of commutative algebra as well. In fact, practical algorithms are known and implemented in various computer algebra packages. In this chapter we will take up a notion that is central to many such algorithms: the notion of a Gröbner basis. Gröbner bases have had interesting theoretical as well as computational applications, and there are currently many open problems in the theory.

In brief, a **Gröbner basis** for an ideal I in S is a set of generators for I with an additional property; **Buchberger's algorithm** yields a simple

and effective method for computing Gröbner bases and syzygies. Through the use of Gröbner bases, many questions about ideals in polynomial rings can be reduced to questions about monomial ideals, which are far easier. The kinds of problems that can be attacked with Gröbner bases can be very roughly divided into two groups: constructive module theory and elimination theory.

Constructive Module Theory

In this heading we group all the operations carried out on modules over a fixed ring. For example, this group includes

- **Perform Division with remainder and compute ideal membership:** Given generators for an ideal $I \subset S$, determine a vector space basis for S/I, and given a polynomial f, compute its image in S/I in terms of this basis. If $f \in I$ (that is, if the image is 0), compute an expression for f as a linear combination of the generators of I.
- **Compute syzygies;** that is, compute the kernel of a map $\varphi : G \to F$ of free S-modules. Equivalently, solve a system of linear equations over S.
- **Compute the intersection of two ideals.**
- **Compute the annihilator of a module.**
- **For ideals I, $J \subset S$, compute the saturation $(I:J^\infty)$.**
- **Compute the module of homomorphisms between two given modules; more generally, compute Ext and Tor.**
- **Compute the Hilbert function and polynomial of a graded module.**

Elimination Theory

In this heading we group the operations that involve two different rings. The most basic operation in this class is

- **Elimination:** Compute the intersection J of an ideal $I \subset k[x_1, \ldots, x_r]$ with a subring $R' = k[x_1, \ldots, x_s]$.

The geometric meaning of elimination is projection: Given an algebraic variety $X \subset \mathbf{A}^r$ defined by the vanishing of the polynomials in I, the projection of X to $\mathbf{A}^s$ is a set whose closure (in the Zariski topology) is defined by J. One of the main uses of elimination is in actually finding solutions for a system of polynomial equations—that is, finding points of a variety. The idea is to reduce the problem to a problem in fewer variables, and eventually to a problem in one variable, where other techniques (factorization of polynomials) can be used. In this chapter we will explain how to use elimination to solve problems such as:

- **Compute the equations satisfied by given elements of an affine ring.** Geometrically, compute the closure of the image of an affine or projective variety under a morphism.

In particular:

- **Find a presentation of the blowup algebra and associated graded ring of a ring $R=S/I$, with respect to an ideal $\mathfrak{m}$.**

- **Given a variety $V \subset \mathbf{A}^r$, find equations for its closure in $\mathbf{P}^r$.**

This chapter is somewhat inhomogeneous. The main results, on which the computational uses of Gröbner bases are founded, are proved in the first part, ending with the treatment of syzygies. To borrow a phrase of Sturmfels', these are the "Gröbner Basics." Next are collected some historical remarks. Subsequent sections on flat families and generic initial ideals present more advanced topics. Some of the ways of applying Gröbner basis techniques in constructive module theory are then described. The novice might want to read just the "Basics" and browse a little among the applications to get a flavor of what is possible. For those wishing to go deeper into the use of computers in commutative algebra and algebraic geometry, I have provided some computer algebra projects, with suggestions for implementation, in addition to more traditional exercises.

Another part of constructive commutative algebra that certainly deserves mention, but that we will not treat here, concerns methods for factoring polynomials. This field is dominated by ideas of E. Berlekamp; a beautiful exposition may be found in the book of Knuth [1969, Vol. II, Section 4.6].

15.1 Monomials and Terms

Since the main idea in the use of Gröbner bases is to reduce all questions to questions about monomials, we begin with these. We write monomials in S using multiindices: If $a = (a_1, \ldots, a_r)$, then x^a will denote the monomial

$$x_1^{a_1} \cdot \ldots \cdot x_r^{a_r}.$$

An ideal generated by such monomials will be called a **monomial ideal.** More generally, let F be a finitely generated free module with basis $\{e_i\}$. **A monomial in F** is an element of the form $m = x^a e_i$ for some i. We will say that such an m **involves the basis element e_i.** A **monomial submodule of F** is a submodule generated by elements of this form. Any monomial submodule M of F may be written as

$$M = \oplus I_j e_j \subset \oplus S e_j = F$$

with I_j the monomial ideal generated by those monomials m such that $me_j \in M$.

A **term** in F is a monomial multiplied by a scalar. Since the monomials form a vector space basis for F, every element $f \in F$ is uniquely expressible as a finite sum of nonzero terms involving distinct monomials, which we call the terms of f; the monomials in these terms will be called the monomials of f. Since we have assumed that k is a field, the distinction between terms and monomials will not play much of a role in our theory.

These definitions all depend on the chosen basis $\{e_i\}$ of F. Whenever possible, we will suppress the actual basis $\{e_i\}$ from our notation, and speak simply of F as a **free module with basis.**

If m, n are monomials of $S, u, v \in k$, and $v \neq 0$, then we say that the term ume_i is **divisible** by the term vne_j if $i = j$ and m is divisible by n in S; the **quotient** is then $um/vn \in S$.

A number of operations are far simpler for monomials than for arbitrary polynomials. For example, the **greatest common divisor** and **least common multiple** of two monomials in S are obtained componentwise: If $b = (b_1, \ldots, b_r)$, then

$$\begin{aligned}\mathbf{GCD}(x^a, x^b) &= x_1^{\min(a_1,b_1)} x_2^{\min(a_2,b_2)} \cdots x_r^{\min(a_r,b_r)},\\ \mathbf{LCM}(x^a, x^b) &= x_1^{\max(a_1,b_1)} x_2^{\max(a_2,b_2)} \cdots x_r^{\max(a_r,b_r)}.\end{aligned}$$

We extend these operations to terms in any free module with basis F: If m, $n \in F$ are terms involving the same basis element e_i of F, then the GCD of m and n will be taken to be the largest monomial in F by which both m and n can be divided. It is easy to write down the intersection or quotient of monomial submodules in terms of these operations; see Exercises 15.3 and 15.7.

If $M \subset F$ is a submodule generated by monomials $m_1, \ldots, m_t$, it is trivial to decide whether a monomial m belongs to M: It does iff it is divisible by at least one of the m_i. More generally, the "membership problem" is easy to solve for a monomial submodule: An arbitrary element $f \in F$ belongs to M iff each of its monomials belongs to M.

Given any set of monomial generators for M, we may remove any that are divisible by others in the set and still have a set of generators for M. In this way we get the unique minimal set of monomials generating M: the set of monomials in M that are minimal elements in the partial order by divisibility on the monomials of F. We will refer to the monomials in this set as **minimal generators of M**.

15.1.1 Hilbert Function and Polynomial

These simple ideas already suffice to compute the Hilbert function and polynomial of a monomial submodule $M \subset F$, or equivalently of the quotient $P = F/M$, quite efficiently. Because the submodule M is a direct sum of modules of the form $I_j e_j$, where the e_j are basis elements of F, we get

$P \cong \oplus S/I_j$. Since the Hilbert function is additive, it suffices to treat the case $P = S/I$, where I is a monomial ideal.

We make an induction using the following idea: Choosing one of the minimal generators n of I, we write $I = (I', n)$, where I' is a monomial ideal generated by fewer monomials than I, and we let d be the degree of n. There is an exact sequence of graded modules and degree 0 maps

$$S(-d) \xrightarrow{\varphi} S/I' \to S/I \to 0,$$

where $S(-d)$ is the free module with generator in degree d and φ is the map that sends the generator of $S(-d)$ to the class of n in S/I'. The kernel of φ is easy to compute. It is the monomial ideal

$$J := (I' : n) = \{m \in S | mn \in I'\},$$

shifted in degree to be a submodule of $S(-d)$. If $I' = (m_1, \ldots, m_t)$ then by Exercise 15.3,

$$J = (m_1/\mathrm{GCD}(m_1, n), \ldots, m_t/\mathrm{GCD}(m_t, n)),$$

so like I', the ideal J has fewer minimal generators than I, and we can suppose by induction that we know the Hilbert function and polynomial of S/I' and S/J.

From the short exact sequence of graded modules

$$0 \to (S/J)(-d) \to S/I' \to S/I \to 0,$$

we get for each integer ν a short exact sequence of vector spaces

$$0 \to (S/J)_{\nu-d} \to (S/I')_\nu \to (S/I)_\nu \to 0.$$

Thus, on the level of Hilbert functions,

$$H_{S/I}(\nu) = H_{S/I'}(\nu) - H_{S/J}(\nu - d),$$

which solves our problem.

By choosing n sensibly, we can make the process much faster: If n contains the largest power of some variable x_1 of any of the minimal generators of I, then the minimal generators of the resulting ideal J will not involve x_1 at all. They will thus involve strictly fewer of the variables than do the minimal generators of I.

This process leads to an expression for the Hilbert function or polynomial as an alternating sum. A variant of the method, which leads directly to an expression for the Hilbert function as a sum of binomial coefficients (all terms positive), is presented in Exercise 15.4. The worst-case behavior of these methods (with the best-known choices for the monomial n) is exponential in the number of variables, and Bayer and Stillman [1992] (from which the preceding method is taken) show that finding the Hilbert function of a monomial ideal is an NP-hard problem in a suitable sense. Nevertheless, in many cases of interest, the method works quite quickly.

15.1.2 Syzygies of Monomial Submodules

Syzygies of monomial submodules are also quite simple. The following result not only gives generators for the syzygies of a monomial submodule, but also gives precise information on the coefficients necessary to express arbitrary syzygies in terms of the generators.

In the following, we let F be a free module with basis and let M be a submodule of F generated by monomials $m_1, \ldots, m_t$. Let

$$\varphi : \oplus_{j=1}^{t} S\varepsilon_j \to F; \qquad \varphi(\varepsilon_j) = m_j$$

be a homomorphism from a free module whose image is M. For each pair of indices i, j such that m_i and m_j involve the same basis element of F, we define

$$m_{ij} = m_i / \mathrm{GCD}(m_i, m_j),$$

and we define σ_{ij} to be the element of $\ker \varphi$ given by

$$\sigma_{ij} = m_{ji}\varepsilon_i - m_{ij}\varepsilon_j.$$

Lemma 15.1. *With notation as above,* $\ker \varphi$ *is generated by the* σ_{ij}.

Proof. We first observe that as a vector space over k, $\ker \varphi$ is the direct sum, over all monomials $n \in F$, of the vector spaces

$$(\ker \varphi)_n = \{\sum a_v n_v \varepsilon_v \in \ker \varphi \mid m_v \text{ divides } n, n_v = n/m_v, \text{ and } a_v \in k\}.$$

Indeed, suppose that

$$\sigma = \sum p_i \varepsilon_i \in S^t, \qquad p_i \in S$$

is a syzygy, so that $\sum p_i m_i = 0$. For any monomial n of F that occurs in one of the $p_j m_j$, and for each i, let $p_{i,n}$ be the term of p_i (if any) such that $p_{i,n} m_i$ is a scalar times n. We must have $\sum p_{i,n} m_i = 0$, so $\sum p_{i,n} \varepsilon_i \in (\ker \varphi)_n$. The representation is clearly unique.

We may now assume that $\sigma = \sum a_v n_v{}^{\varepsilon} (\ker \varphi)_n$, and we prove that σ is in the module generated by the σ_{ij} by induction on the number of nonzero terms of σ. If $\sigma \neq 0$, then because σ is a syzygy, at least two of the $a_v n_v$ must be nonzero, say the ith and the jth, with $i < j$. It follows that n is divisible both by m_i and m_j, and thus n_i is divisible by

$$\mathrm{LCM}(m_i, m_j)/m_i = m_j / \mathrm{GCD}(m_i, m_j) = m_{ji}.$$

Consequently, we may subtract a scalar times $(n_i / m_{ji}) \sigma_{ij}$ from σ to get a relation with fewer terms. □

The proof actually gives a stronger result, which we will use in the proof of Theorem 15.8:

Lemma 15.1 bis. *With notation as in Lemma* 15.1, *every element of* $\ker\varphi$ *is uniquely expressible as a sum of elements* $\tau = \sum a_v n_v \varepsilon_v \in \ker\varphi$ *such that all the* $n_v m_v$ *are equal to the same monomial* $n \in F$. *For such an element we may write*

$$\tau = \sum n_{ij}\sigma_{ij},$$

where the sum is over all $i < j$ *such that* $\mathrm{LCM}(m_i, m_j)$ *divides* n, *and where* n_{ij} *is a scalar times the monomial* $n/\mathrm{LCM}(m_i, m_j) = n_i/m_{ji}$.

Proof. The first paragraph of the proof of Lemma 15.1 proves the first statement. For the second, look again at the last paragraph of the proof of Lemma 15.1. The element σ_{ij} used there meets the conditions of Lemma 15.1 bis, and we never introduce any new term in τ in the course of the induction. □

The syzygies σ_{ij} in Lemma 15.1 are sometimes called **divided Koszul relations** because of their similarity to the relations in the Koszul complex that we shall study in Chapter 17. We have shown that they generate all the syzygies on monomial ideals, but in general they do not form a minimal set of generators (see Exercise 15.6). In Exercise 17.11 we will see that a very similar construction gives a whole (nonminimal) free resolution of a monomial submodule, which is a kind of "divided Koszul complex."

15.2 Monomial Orders

If $J \subset S$ is a monomial ideal, then the set B of all monomials not in J forms a vector space basis for S/J that makes computation in S/J quite convenient. If I is an arbitrary ideal of S, we would like to obtain a similarly simple picture of S/I. Since the monomials of S form a vector space basis, their images span S/I, and a maximal linearly independent subset B will be a basis. These exist by Zorn's lemma, so any S/I has such a **monomial basis**.

If we can choose B to be the complement of the set of monomials in a monomial ideal J, as in the case where I is itself a monomial ideal, we get an extra advantage. Because a monomial ideal can be specified by giving finitely many monomial generators, it is easy to determine whether a given monomial is in B: We must simply test for divisibility by one of the generators of J. We will show in Theorem 15.3 that there is a monomial basis B for any S/I obtained in this way. We begin here with some remarks to motivate the construction.

First, if J is a monomial ideal and B is the set of monomials not in J, then it is not hard to see that the elements of B remain linearly independent modulo an ideal I iff $(*)$ J contains at least one monomial from every polynomial in I. For the set B to be a basis of S/I, the ideal J must (at least!) be minimal with property $(*)$.

As a first example, let $I = (m_1 + m_2)$ be a principal ideal generated by the sum of two monomials m_i. A monomial ideal contains at least one monomial from each polynomial in I iff it contains one of the monomial ideals (m_i). However, if m_1 divides m_2 and we take $J = (m_1)$, then B will not be a basis since J is not minimal: m_1 itself is superfluous. Taking $J = (m_2)$ in these circumstances *does* make B a basis, however; we will prove a much more general statement in a moment, but the reader may wish to pause to think through this special case.

To find a monomial ideal J that contains at least one monomial from each polynomial of I it seems natural to look for a method of choosing one monomial from each polynomial of S. Given such a method, we can apply it to choose a monomial from each polynomial in I, and use the chosen monomials to generate J. To make J minimal, some interesting additional conditions must be met.

Suppose for example that m_1, m_2, m_3 are distinct monomials of the same degree d and that

$$I = (m_1 + m_2, m_2 + m_3) + (\text{all monomials of degree} > d).$$

Suppose that we have chosen m_1 from $m_1 + m_2$ and m_2 from $m_2 + m_3$ to put into J. The ideal I also contains

$$(m_1 + m_2) - (m_2 + m_3) = m_1 - m_3.$$

We must at this point choose m_1 (rather than m_3) to put into J, because if we put m_3 into J, then J would not be minimal. Thus if we write $m_1 > m_2$ for the relation "m_1 is chosen over m_2", then $>$ must satisfy the axiom for an order relation, $m_1 > m_2 > m_3 \Rightarrow m_1 > m_3$. A more careful analysis shows that the same thing is true even when the m_i have different degrees.

Thus we must totally order the monomials of S, and put into J the greatest monomial in each polynomial of I. Because we wish to take J to be an ideal, there are two further requirements that the order $>$ must satisfy with respect to multiplication.

First, as shown in the first example, $>$ must refine the partial order defined by divisibility: That is, if m_2 is divisible by m_1, we must take $m_2 > m_1$.

Second, $>$ must be preserved by multiplication: Suppose that $I = (m_1 + m_2)$ and that we have chosen $m_1 > m_2$ so that $m_1 \in J$ and m_1 does not divide m_2. Then $nm_1 + nm_2 \in I$, but already, since J is an ideal, $nm_1 \in J$, so choosing $nm_2 > nm_1$ would lead (under many circumstances) to nonmimimal sets J. Thus we must have $nm_1 > nm_2$. The following definition encapsulates these conditions.

Definition. *Let F be a free S-module with basis. A* ***monomial order*** *on F is a total order $>$ on the monomials of F such that if m_1, m_2 are monomials of F and $n \neq 1$ is a monomial of S, then*

$$m_1 > m_2 \textit{ implies } nm_1 > nm_2 > m_2.$$

Bearing in mind that we are supposing F to be finitely generated, the second inequality has an extremely useful consequence.

Lemma 15.2. *Let F be a free S-module with basis. Any monomial order on F is **Artinian** (every subset has a least element).*

Proof. If X is a set of monomials of F, then since S is Noetherian the submodule of F generated by X is already generated by a finite subset $Y \subset X$. The least element of Y will be the least element in X because every element of X is a multiple, by a monomial in S, of an element of Y. □

We will extend this notation to terms: If um and vn are terms with $0 \neq u, v \in k$, and m, n are monomials with $m > n$ (respectively, $m \geq n$) then we say $um > vn$ (respectively, $um \geq vn$). Note that this is *not* a partial order on terms, since even if $u \neq v$ we have $um \geq vm$ and $vm \geq um$. It is nonetheless convenient.

If $>$ is a monomial order, then for any $f \in F$ we define the **initial term of f**, written $\mathbf{in}_>(f)$ to be the greatest term of f with respect to the order $>$, and if M is a submodule of F we define $\text{in}_>(M)$ to be the monomial submodule generated by the elements $\text{in}_>(f)$ for all $f \in M$. When there is no danger of confusion we will simply write **in** in place of $\mathbf{in}_>$.

Note that if $p \in S$ and $f \in F$ and we write n for the (unique) term of p such that $n \text{ in}(f)$ is greatest, then $\text{in}(pf) = n \text{ in}(f)$. If m is a term of f other than $\text{in}(f)$ and n' is a term of p other than n, we will have

$$\begin{aligned} n \text{ in}(f) &> n' \text{ in}(f) && \text{(by hypothesis)} \\ &> n'm && \text{(because $>$ is a monomial order).} \end{aligned}$$

Monomial orders do all that we might have hoped

Theorem 15.3 (Macaulay). *Let F be a free S-module with basis, and let M be an arbitrary submodule. For any monomial order $>$ on F, the set B of all monomials not in $\text{in}_>(M)$ forms a basis for F/M.*

Proof. To show that B is linearly independent, note that if there were a dependence relation

$$p = \sum u_i m_i \in M \qquad m_i \in B, \qquad 0 \neq u_i \in k$$

then $\text{in}(p) \in \text{in}(M)$. Since $\text{in}(p)$ is one of the m_i, which are supposed to be in B, this is a contradiction.

Now suppose that B does not span F/M. Among the set of elements of F that are not in the span of M and B, we may take f to be one with minimal initial term. If $\text{in}(f)$ were in B, we could subtract it, getting a polynomial with a still smaller initial term. Thus we may suppose that $\text{in}(f) \in \text{in}(M)$. Subtracting an element of M with the same initial term as f results in a similar contradiction. □

Monomial orders abound. Here are some significant examples with $F = S$. We write $a = (a_1, \ldots, a_r)$ and $b = (b_1, \ldots, b_r)$ for multiindices, and set $m = x^a$, and $n = x^b$. By renaming the variables, we may always achieve $x_1 > x_2 > \ldots > x_r$, and we will only describe orders with this property.

Lexicographic order. *$m >_{\text{lex}} n$ iff $a_i > b_i$ for the **first** index i with $a_i \neq b_i$.*

Homogeneous lexicographic order. *$m >_{\text{hlex}} n$ iff $\deg m > \deg n$ or $\deg m = \deg n$ and $a_i > b_i$ for the **first** index i with $a_i \neq b_i$.*

If we are given a sequence of partial orders $>_1, >_2, \ldots$, then we may define the partial order that is their **lexicographic product** to be the order in which $m > n$ if $m >_i n$ for the first i such that m and n are comparable with respect to the order $>_i$. We sometimes say that the lexicographic product order is the order $>_1$ **refined by** the order $>_2$ refined by $\ldots$. The homogeneous lexicographic order is the lexicographic product of the partial order by degree ($m > n$ if deg $m >$ deg n) refined by the partial orders by the degree in x_1, the degree in $x_2, \ldots$.

If $r = 1$ then the requirement that $nm_2 > m_2$ for a monomial n not equal to 1 shows that there is a unique monomial order on S: the order by degree. Similarly, if $r = 2$, then there is only one monomial order on S that refines the order by degree and satisfies our convention $x_1 > x_2$. To see this, suppose $m = x_1^{a_1} x_2^{a_2}$ and $n = x_1^{b_1} x_2^{b_2}$ have the same degree $a_1 + a_2 = b_1 + b_2$. If $a_1 > b_1$, so $\varepsilon := a_1 - b_1 > 0$, then writing $p = x_1^{b_1} x_2^{a_2}$ for the greatest common divisor gives

$$m = x_1^\varepsilon p,$$
$$n = x_2^\varepsilon p.$$

But $x_1 > x_2$ implies $x_1^\varepsilon > x_2^\varepsilon$ (in fact induction gives $x_1^\varepsilon > x_1^{\varepsilon-1} x_2 > x_2^\varepsilon$), so $m > n$.

There are in general many other orders. By far the most important is the following.

Reverse lexicographic order. *$m >_{\text{rlex}} n$ iff $\deg m > \deg n$ or $\deg m = \deg n$ and $a_i < b_i$ for the **last** index i with $a_i \neq b_i$.*

Note the direction of the inequality $\boldsymbol{a_i < b_i}$. The name "reverse lexicographic" comes from the fact that, on the monomials of a given degree, this is the reverse of the order obtained by reversing the order of the variables and using homogeneous lexicographic order. (The opposite of lexicographic order is itself not an order in our sense; can the reader see why?) Reverse lexicographic order was introduced by Macaulay [1927]. The difference between the homogeneous lexicographic and reverse lexicographic

orders is subtle, but the use of reverse lexicographic order in place of homogeneous lexicographic order in the algorithms described next sometimes improves the efficiency of computation enormously (Bayer and Stillman [1987a and b]). See the section on generic coordinates for a hint of a possible reason.

The first case in which $>_{\text{hlex}}$ and $>_{\text{rlex}}$ could differ is for quadratic monomials in three variables. Here indeed we have

$$x_1x_3 >_{\text{hlex}} x_2^2$$

while

$$x_1x_3 <_{\text{rlex}} x_2^2.$$

Roughly, we can describe the difference by saying that if m and n have the same degree, then $m >_{\text{hlex}} n$ iff m involves **more** from the **beginning** of the list of variables, while $m >_{\text{rlex}} n$ iff m involves **less** from the **end** of the list of variables. Most of the uses made of these orders depend on the following easily verified properties (which actually characterize them; see Exercise 15.10). These properties make it clear that the subtlety above is the difference between a subring and an ideal.

Proposition 15.4 (Characteristic properties of lex, hlex, and rlex).

a. If $\text{in}_{\text{lex}}(f) \in k[x_s, \ldots, x_r]$ *for some* s, *then* $f \in k[x_s, \ldots, x_r]$.

b. $>_{\text{hlex}}$ *refines the order by total degree; and if* f *is homogeneous with* $\text{in}_{\text{hlex}}(f) \in k[x_s, \ldots, x_r]$ *for some* s, *then* $f \in k[x_s, \ldots, x_r]$.

c. $>_{\text{rlex}}$ *refines the order by total degree; and if* f *is homogeneous with* $\text{in}_{\text{rlex}}(f) \in (x_s, \ldots, x_r)$ *for some* s, *then* $f \in (x_s, \ldots, x_r)$.

Weight orders: We define a **weight function** λ for S to be a linear function $\mathbf{R}^r \to \mathbf{R}$; λ will be called **integral** if it comes from a linear map $\mathbf{Z}^r \to \mathbf{Z}$. Any weight function λ defines a partial order $>_\lambda$, called the **weight order associated to $\boldsymbol{\lambda}$,** by the rule $m = x^a >_\lambda n = x^b$ iff $\lambda(a) > \lambda(b)$. We say that λ is **compatible** with a given monomial order $>$ if $m >_\lambda n$ implies $m > n$. Similar things could be done for a free module, but we shall not use this. There are always compatible weight orders: In fact, it can be shown (Robbiano [1986]; and see Exercises 15.11–15.13) that every monomial order is the lexicographic product of r weight orders, the first of which is necessarily compatible with the given order. For example, defining $\pi_i : \mathbf{R}^r \to \mathbf{R}$ to be the projection onto the ith coordinate, the lexicographic order is the lexicographic product of the weight orders corresponding to the π_i, while the reverse lexicographic order is the lexicographic product of the weight orders corresponding to the total degree function $\sigma = \Sigma\pi_i$ and the functions $-\pi_r, -\pi_{r-1}, \ldots, -\pi_1$ (the last of which may of course be omitted). In fact, any monomial order $>$ can be approximated by a single weight order

$>_\lambda$ in the sense that $>_\lambda$ can be made to agree with $>$ on any given finite set of pairs of monomials; see Exercise 15.12.

In Proposition 15.16 we will have occasion to use a small extension of the notion of initial term: Given a weight order λ on the polynomial ring S, we define $\text{in}_\lambda(f)$ to be the sum of all those terms of f that are maximal for $>_\lambda$.

One way of getting monomial orders on a free module F with given basis $\{e_i\}$ is to choose a monomial order $>$ on S, choose an order $\succ$ among the e_i, and use a lexicographic product of the partial orders on the monomials of F induced by $>$ and $\succ$. In particular, **a reverse lexicographic order on *F*** is the result of refining the reverse lexicographic order on the monomials of S by an order of the basis e_i in this way.

Now let F be a free S-module with basis and let M be a submodule of F. It turns out to be extremely useful to know the modules $\text{in}_>(M)$ with respect to various orders $>$. "Knowing" such a module means of course having a system of generators for it. It turns out to be practical to ask for a little more information: a system of generators for $\text{in}_>(M)$, and for each one an element of M whose initial form it is. The following central definition encapsulates a convenient description of this information.

Definition. *A* ***Gröbner basis*** *with respect to an order $>$ on a free module with basis F is a set of elements $g_1, \ldots, g_t \in F$ such that if M is the submodule of F generated by $g_1, \ldots, g_t$, then $\text{in}_>(g_1), \ldots, \text{in}_>(g_t)$ generate $\text{in}_>(M)$. We then say that $g_1, \ldots, g_t$ is a* ***Gröbner basis for M****.*

Examples. The case of no variables: Let S be a field and let F be a vector space of dimension s, with basis $\{e_i\}$; we may identify elements of F with column vectors of length s. The only monomials of F are the e_i; let $>$ be the monomial order in which $e_1 > e_2 > \ldots$.

A set of elements $g_1, \ldots, g_t \in F$ is simply an $s \times t$ matrix G over S. The set G is a Gröbner basis iff it contains a maximal linearly independent set in "echelon form," that is, if some maximal independent subset of the column vectors g_i have their first nonzero entries in distinct rows, as in Figure 15.1.

The case of one variable: Next consider the case $S = k[x]$, a polynomial ring over k in one variable, and take $F = S$. The only monomial order is the order by degree. A submodule $M \subset F$ is then just an ideal. The monomial ideal $\text{in}(M)$ is generated by x^d where d is the smallest degree of any polynomial in M; thus a Gröbner basis of M consists of any set of generators of M containing an element of minimal degree. Note that Lemma 15.5 provides a proof that an ideal is generated by any element of minimal degree.

There is a Gröbner basis for any submodule M of F, with respect to any monomial order: If $g_1, \ldots, g_t$ are generators for M that are not a Gröbner

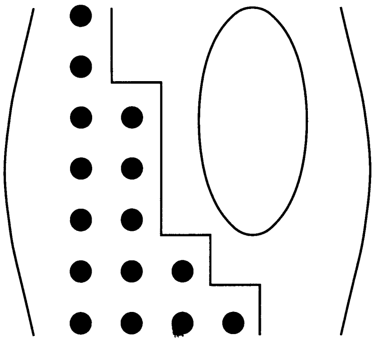


FIGURE 15.1.

basis, then to get a Gröbner basis we simply adjoin elements $g_{t+1}, \dots, g_{t'}$ of M until the initial terms $\mathrm{in}(g_1), \dots, \mathrm{in}(g_{t'})$ generate $\mathrm{in}(M)$. This is possible by the Hilbert basis theorem. The following lemma shows that any set of elements of M whose initial terms generate $\mathrm{in}(M)$ actually generate M. Thus, to check that a set of elements is a Gröbner basis for M, it is enough to check that their initial terms generate $\mathrm{in}(M)$.

Lemma 15.5. *If $N \subset M \subset F$ are submodules and $\mathrm{in}(N) = \mathrm{in}(M)$ with respect to a monomial order, then $N = M$.*

Proof. If $N \neq M$, then there would be an element $f \in M$ not in N whose initial term is smallest among initial terms of elements not in N. Since $\mathrm{in}(f) \in \mathrm{in}(M) = \mathrm{in}(N)$, we may write $\mathrm{in}(f) = \mathrm{in}(g)$ with $g \in N$. But then $f - g \in M$, $f - g \notin N$, and $f - g$ and has smaller initial term than f —a contradiction. □

Once we can compute Gröbner bases, Lemma 15.5 suffices to solve the "submodule membership" problem: Given a submodule M of a free module with basis F and an element $f \in F$, decide whether $f \in M$. To do this, choose a monomial order on F and find $\mathrm{in}(M)$ and $\mathrm{in}(M + Sf)$. By Lemma 15.5, the element f is in M iff $\mathrm{in}(M) = \mathrm{in}(M + Sf)$. This is easy to test because $\mathrm{in}(M)$ and $\mathrm{in}(M + Sf)$ are monomial submodules. (In practice one would probably use the division algorithm presented in the next section instead of this method.)

A Gröbner basis $g_1, \dots, g_t$ such that $\mathrm{in}(g_i)$ does not divide $\mathrm{in}(g_j)$ for any $i \neq j$ (that is, such that the $\mathrm{in}(g_i)$ are a minimal set of generators for the monomial submodule they generate) is called a **minimal** Gröbner basis. We can make a Gröbner basis for M into a minimal Gröbner basis just by leaving out some elements. More interesting perhaps, a Gröbner basis $g_1, \dots, g_t$ such that $\mathrm{in}(g_i)$ does not divide any term of g_j for $i \neq j$ is said to be **reduced**. If we assume in addition that $\mathrm{in}(g_i)$ is a monomial (that is, the coefficient from k is 1), then we get something uniquely defined in terms of the submodule, the basis of F, the choice of variables in S, and the choice of order—see Exercise 15.14.

15.3 The Division Algorithm

One of the most elementary and useful operations with polynomials in one variable is "division with remainder"; given polynomials f and g this algorithm constructs an expression of the form $f = f_1 g + f'$ with $\deg f_1 g = \deg f$, and $\deg f' < \deg g$ (or possibly $f' = 0$). Given such an expression, f' is called the *remainder on division.* If we order the monomials of $S = k[x_1]$ by degree (that is: $x_1^s < x_1^t$ iff $s < t$), then we can restate the conditions on f_1 and f' above by saying that f' has no monomials in the initial ideal of (g), and $\mathrm{in}(f) \geq \mathrm{in}(f_1 g)$ (actually the terms in question are equal). We will now extend this process to the general case. A side effect will be an algorithm for computing a Gröbner basis.

Proposition–Definition 15.6. *Let F be a free S-module with basis and monomial order $>$. If $f, g_1, \ldots, g_t \in F$ then there is an expression*

$$f = \sum f_i g_i + f' \quad \text{with } f' \in F, \qquad f_i \in S,$$

where none of the monomials of f' is in $(\mathrm{in}(g_1), \ldots, \mathrm{in}(g_t))$ and

$$\mathrm{in}(f) \geq \mathrm{in}(f_i g_i)$$

for every i. Any such f' is called a ***remainder*** *of f with respect to $g_1, \ldots, g_t$, and an expression $f = \sum f_i g_i + f'$ satisfying the condition of the proposition is called a* ***standard expression*** *for f in terms of the g_i.*

The proof consists of an algorithm for finding a standard expression of the desired sort.

Division Algorithm 15.7. *Let F be a free S-module with basis and a fixed monomial order. If $f, g_1, \ldots, g_t \in F$, then we may produce a standard expression*

$$f = \sum m_u g_{s_u} + f'$$

for f with respect to $g_1, \ldots, g_t$ by defining the indices s_u and the terms m_u inductively. Having chosen $s_1, \ldots, s_p$ and $m_1, \ldots, m_p$, if

$$f'_p := f - \sum_{u=1}^{p} m_u g_{s_u} \neq 0$$

and m is the maximal term of f'_p that is divisible by some $\mathrm{in}(g_i)$, then we choose

$$s_{p+1} = i,$$
$$m_{p+1} = m / \mathrm{in}(g_i)$$

This process terminates when either $f'_p = 0$ or no $\text{in}(g_i)$ *divides a monomial of f'_p; the remainder f' is then the last f'_p produced.*

Lemma 15.2 guarantees that the algorithm terminates after finitely many steps because the maximal term of f'_p divisible by some g_i decreases at each step.

The division algorithm is most important in the case when the g_i form a Gröbner basis for a submodule M of F; then from the conditions of a standard expression, we see that the remainder f' gives the expression for f mod M in terms of the basis of F/M guaranteed by Theorem 15.3. In fact, since Gröbner bases always exist, the division algorithm gives us another, more constructive, version of the second half of the proof of Theorem 15.3.

Note that standard expressions are far from unique: The division algorithm as we have stated it is indeterminate, in that the remainder depends on some choices made in carrying out the process. This is occasionally useful in that some choices are more efficient than others. In fact, the division algorithm still terminates if at each stage we simply choose *some* term of f'_p divisible by some $\text{in}(g_i)$, instead of the greatest term. This gives a still-more indeterminate version of the division algorithm, which works just as well for the purposes of this chapter (see Exercise 15.16).

It is sometimes useful to have a **determinate division algorithm**; we can do this by specifying (for example) that at each step we take m to be the greatest monomial of f'_p that is divisible by some $\text{in}(g_i)$, and s_{p+1} the smallest i for which this division is possible. Such determinate division gives a unique standard expression satisfying certain auxiliary conditions (see Exercise 15.17).

It is easy to check that the division algorithm works just as well for weight orders and other monomial partial orders, so long as the initial forms of all the polynomials considered are terms (the initial form is by definition the sum of *all* the maximal terms).

15.4 Gröbner Bases

The division algorithm leads to a computation of Gröbner bases and syzygies on them, the two major topics of this chapter. We will make use of the following notation:

Let F be a free module over S with basis and monomial order $>$. Let $g_1, \dots, g_t$ be nonzero elements of F. Let $\oplus S\varepsilon_i$ be a free module with basis $\{\varepsilon_i\}$ corresponding to the elements $\{g_i\}$ of F, and let

$$\varphi : \oplus S\varepsilon_i \to F; \qquad \varepsilon_i \mapsto g_i$$

be the corresponding map.

For each pair of indices i, j such that $\mathrm{in}(g_i)$ *and* $\mathrm{in}(g_j)$ *involve the same basis element of F, we define*

$$m_{ij} = \mathrm{in}(g_i)/\,GCD\ (\mathrm{in}(g_i), \mathrm{in}(g_j)) \in S,$$

and we set

$$\sigma_{ij} = m_{ji}\varepsilon_i - m_{ij}\varepsilon_j,$$

so that the σ_{ij} generate the syzygies on the elements $\mathrm{in}(g_i)$ *by Lemma* 15.1. *For each such pair i, j, we choose a standard expression*

$$m_{ji}g_i - m_{ij}g_j = \sum f_u^{(ij)} g_u + h_{ij}$$

for $m_{ji}g_i - m_{ij}g_j$ with respect to $g_1, \ldots, g_t$. Note that $\mathrm{in}(f_u^{(ij)} g_u) < \mathrm{in}(m_{ji}g_i)$. *For convenience we set $h_{ij} = 0$ if* $\mathrm{in}(g_i)$ *and* $\mathrm{in}(g_j)$ *involve distinct basis elements of F.*

With this notation we have:

Theorem 15.8 (Buchberger's Criterion). *The elements $g_1, \ldots, g_t$ form a Gröbner basis iff $h_{ij} = 0$ for all i and j.*

Proof. Let $M = (g_1, \ldots, g_t) \in F$. From the expression for h_{ij} we see that $h_{ij} \in M$, and thus $\mathrm{in}(h_{ij}) \in \mathrm{in}(M)$. If $g_1, \ldots, g_t$ is a Gröbner basis, then as remarked just after the proof of the division algorithm, the definition of a standard expression shows that $h_{ij} = 0$.

Conversely, suppose that all $h_{ij} = 0$, so that $\varphi(\sigma_{ij}) = \sum f_u^{(ij)} g_u$ with $\mathrm{in}(f_u^{(ij)} g_u) < \mathrm{in}(m_{ji}g_i)$. If $g_1, \ldots, g_t$ is not a Gröbner basis, then we may choose an expression

$$f = \sum_u f_u g_u \quad \text{with} \quad \mathrm{in}(f) \notin (\mathrm{in}(g_1), \ldots, \mathrm{in}(g_t)).$$

Let m be the maximum among the terms $\mathrm{in}(f_u g_u)$. We may suppose that the expression for f has been chosen so that m is as small as possible. Now let $\sum' f_v g_v$ be the sum of all those $f_v g_v$ for which $\mathrm{in}(f_v g_v)$ is m times a scalar. We may write $\mathrm{in}(f_v g_v) = n_v\ \mathrm{in}(g_v)$ for some term n_v of f_v. If the sum of the corresponding initial terms $\sum' n_v \mathrm{in}(g_v)$ is nonzero, then it is the initial term of f; as it is a multiple of m, it is a multiple of $\mathrm{in}(g_v)$, contradicting the choice of f. Thus

$$\sum{}' n_v\, \mathrm{in}(g_v) = 0,$$

so that $\sum' n_v \varepsilon_v$ is a syzygy among the $\mathrm{in}(g_v)$.

By Lemma 15.1 bis, we may write $\sum' n_v \varepsilon_v = \sum_{i<j} a_{ij}\sigma_{ij}$ where a_{ij} is a scalar times $m/\,\mathrm{in}(g_i)$. If we apply φ and substitute $\sum f_u^{(ij)} g_u$ for $\varphi(\sigma_{ij})$, we find a relation of the form

$$\sum{}' n_v g_v = \sum h_s g_s$$

with all $\mathrm{in}(h_s g_s) < m$. Subtracting the expression $\sum' n_v g_v - \sum h_s g_s$ from the expression for f and cancelling the terms of $\sum' n_v in(g_v)$, we get a new expression for f of the same form but where the maximum of the $\mathrm{in}(f_u g_u)$ is smaller, contradicting our construction. □

One can slightly sharpen the criterion in a way that is occasionally useful in practice: It is enough for h_{ij} to be zero for any subset of pairs i, j such that the corresponding σ_{ij} generate all the syzygies on the elements $\mathrm{in}(g_i)$, Also, if $F = S$ we may further omit any pair i, j such that $\mathrm{GCD}(\mathrm{in}(g_i), \mathrm{in}(g_j)) = 1$; see Exercises 15.19 and 15.20.

From Theorem 15.8 we get an effective method for computing Gröbner bases and syzygies.

Buchberger's Algorithm 15.9: *In the situation of Theorem* 15.8, *suppose that M is a submodule of F, and let $g_1, \dots, g_t \in M$ be a set of generators of M. Compute the remainders h_{ij}. If all the $h_{ij} = 0$, then the g_i form a Gröbner basis for M. If some $h_{ij} \neq 0$, then replace $g_1, \dots, g_t$ with $g_1, \dots, g_t, h_{ij}$, and repeat the process. As the submodule generated by the initial forms of $g_1, \dots, g_t, h_{ij}$ is strictly larger than that generated by the initial forms of $g_1, \dots, g_t$, this process must terminate after finitely many steps.*

The process involved in Buchberger's algorithm is even more useful than first appears: Theorem 15.10 shows that the equations $h_{ij} = 0$ that result if the g_i are a Gröbner basis give all the syzygies on M (this is Schreyer's algorithm for computing syzygies). A worked example is given at the end of the following subsection.

There is a fairly sharp "worst-case" upper bound b for the degree of the elements of the Gröbner basis for a homogeneous ideal $(g_1, \dots, g_t) \subset S$ (the nonhomogeneous case can be reduced to this) with respect to the lexicographic order. The bound, which is due to Möller and Mora [1984], is in terms of:

$r =$ the number of variables,
$d =$ the maximum degree of the polynomials g_i, and
$s =$ the degree of the Hilbert polynomial (this is one less than the dimension; it is between 0 and $r - 1$).

The bound is

$$b = ((r+1)(d+1)+1)^{2^{(s+1)}(r+1)},$$

and thus is potentially doubly exponential in the number of variables.

For example, for the homogeneous ideal of a curve of degree δ in $\mathbf{P}^3$, it is known that we can take $r = 4$, $s = 2$, $d = \delta - 1$ and get

$$b \sim (5\delta + 1)^{32}.$$

This estimate is so large as to suggest that Buchberger's algorithm and Gröbner bases would be useless in practice. Fortunately, this is not at all the case: In actual use the algorithm terminates quite quickly on very many problems of interest. There is a partial understanding of why this is so, and various other bounds are known in some special cases; see for example Gruson, Lazarsfeld, and Peskine [1983], Winkler [1984], and Bayer and Stillman [1987a].

15.5 Syzygies

We retain the notation introduced in the previous section.

There is a bonus from Buchberger's algorithm: an effective method for computing syzygies. The process in Algorithm 15.9 gives a linear combination of the g_u that is equal to h_{ij}. Thus if $h_{ij} = 0$ we get a linear relation among the g_u—that is, a syzygy. It turns out that these syzygies generate the entire module of syzygies on the g_i.

We retain the notation developed for Theorem 15.8. In addition, for $i < j$ such that $\mathrm{in}(g_i)$ and $\mathrm{in}(g_j)$ involve the same basis element of F, we set

$$\tau_{ij} = m_{ji}\varepsilon_i - m_{ij}\varepsilon_j - \sum_u f_u^{(ij)}\varepsilon_u.$$

Theorem 15.10 (Schreyer). *With notation as above, suppose that $g_1, \dots, g_t$ is a Gröbner basis. Let $>$ be the monomial order on $\oplus_{j=1}^t S\varepsilon_j$ defined by taking $m\varepsilon_u > n\varepsilon_v$ iff*

$$\mathrm{in}(mg_u) > \mathrm{in}(ng_v) \quad \textit{with respect to the given order on } F$$

or

$$\mathrm{in}(mg_u) = \mathrm{in}(ng_v) \quad \textit{(up to a scalar)} \quad \textit{but } u < v.$$

The τ_{ij} generate the syzygies on the g_i. In fact, the τ_{ij} are a Gröbner basis for the syzygies with respect to the order $>$, and $\mathrm{in}(\tau_{ij}) = m_{ji}\varepsilon_i$.

Proof. We show first that the initial term of τ_{ij} is $m_{ji}\varepsilon_i$. We have

$$m_{ji}\,\mathrm{in}(g_i) = m_{ij}\,\mathrm{in}(g_j),$$

and these terms are by hypothesis greater than any that appear in the $f_u^{(ij)} g_u$. Thus, $\mathrm{in}(\tau_{ij})$ is either $m_{ji}\varepsilon_i$ or $-m_{ij}\varepsilon_j$ by the first part of the definition of $>$, and since $i < j$ we have $m_{ji}\varepsilon_i > m_{ij}\varepsilon_j$.

Now we show that the τ_{ij} form a Gröbner basis. Let $\tau = \sum f_v\varepsilon_v$ be any syzygy. We must show that $\mathrm{in}(\tau)$ is divisible by one of the $\mathrm{in}(\tau_{ij})$; that is, $\mathrm{in}(\tau)$ is a multiple of some $m_{ji}\varepsilon_i$ with $i < j$.

For each index v, set $n_v\varepsilon_v = \mathrm{in}(f_v\varepsilon_v)$. Since these terms cannot cancel with each other, we have $\mathrm{in}(\sum f_v\varepsilon_v) = n_i\varepsilon_i$ for some i. Let $\sigma = \sum' n_v\varepsilon_v$ be

the sum over all indices v for which $n_v \operatorname{in}(g_v) = n_i \operatorname{in}(g_i)$ up to a scalar; all indices v in this sum must be $\geq i$ because we have assumed that $n_i\varepsilon_i$ is the initial term of τ.

Thus, σ is a syzygy on the $\operatorname{in}(g_v)$ with $v \geq i$. By Lemma 15.1, all such syzygies are generated by the σ_{uv} for $u, v \geq i$, and the ones in which ε_i, appears are the σ_{ij} for $j > i$. It follows that the coefficient n_i is in the ideal generated by the m_{ji} for $j > i$, and we are done. □

As with Buchberger's criterion, we can sharpen this result slightly in a useful way: To find a set of τ_{ij}, which generate all the syzygies on the g_i, it is enough to take a set of pairs i, j such that the σ_{ij} generate the syzygies of the elements $\operatorname{in}(g_i)$. See Exercise 15.18.

If we wish to use Theorem 15.10 to compute syzygies on a fixed set of elements $g_1, \ldots, g_t$, we first use Buchberger's algorithm to obtain a Gröbner basis for $(g_1, \ldots, g_t)$ and the syzygies on the Gröbner basis elements. To get the syzygies on the g_i, we need only substitute into these syzygies the expressions for the Gröbner basis elements in terms of the g_i.

This process usually will not give us a minimal set of syzygies: To replace it with a minimal set (say in the case where everything is homogeneous, so that minimal resolutions are well defined; see Chapter 20) we must do some further work, finding at least the degree $= 0$ syzygies among the nonminimal syzygies, and using them to eliminate superfluous relations. Nevertheless, this process is by far the most efficient method known for computing syzygies.

An example will help to clarify all this.

Example. The simplest nontrivial Gröbner basis computation. Take $g_1 = x^2$, $g_2 = xy + y^2$. We will find a Gröbner basis with respect to the lexicographic order, taking $x > y$. We have $\operatorname{in}(g_1) = x^2$, $\operatorname{in}(g_2) = xy$. The GCD is x. We apply the division algorithm to

$$(\operatorname{in}(g_2)/x)g_1 - (\operatorname{in}(g_1)/x)g_2 = -xy^2.$$

In the first step we add yg_2, getting y^3. Since this is not divisible by either initial form, it is the remainder; as it is not 0, we take it as $g_3 = y^3$, and we have the syzygy

$$\tau_{1,2} : y\varepsilon_1 - x\varepsilon_2 + y\varepsilon_2 - \varepsilon_3.$$

Since g_1 and g_3 are monomials, we get from them a syzygy

$$\tau_{1,3} : y^3\varepsilon_1 - x^2\varepsilon_3.$$

The only other pair to check is g_2 and g_3: Applying the division algorithm to

$$(\operatorname{in}(g_3)/y)g_2 - (\operatorname{in}(g_2)/y)g_3 = y^4,$$

we subtract yg_3 and find a remainder of 0. Thus we get the syzygy

$$\tau_{2,3} : y^2\varepsilon_2 - x\varepsilon_3 - y\varepsilon_3 = y^2\varepsilon_2 - (x + y)\varepsilon_3.$$

From Buchberger's criterion we see now that

$$x^2, xy + y^2, y^3$$

is a Gröbner basis, and from Theorem 15.10 we know that $\tau_{1,2}$, $\tau_{2,3}$, and $\tau_{1,3}$ generate the syzygies on them. If we wish to derive from this a set of generators for the syzygies on the original generators g_1, g_2, we must substitute the expression for g_3 in terms of g_1 and g_2 given by the syzygy $\tau_{1,2}$ into the other syzygies. We get

$$\begin{aligned} \tau_{1,2} &: \ 0 \\ \tau_{1,3} &: \ y^3\varepsilon_1 - x^2(y\varepsilon_1 - x\varepsilon_2 + y\varepsilon_2) = (y^3 - x^2y)\varepsilon_1 + (x^3 - x^2y)\varepsilon_2 \\ \tau_{2,3} &: \ y^2\varepsilon_2 - (x+y)\varepsilon_3 = y^2\varepsilon_2 - (x+y)(y\varepsilon_1 - x\varepsilon_2 + y\varepsilon_2) \\ &\qquad = x^2\varepsilon_2 - (xy + y^2)\varepsilon_1. \end{aligned}$$

We see that $\tau_{1,3} = (x - y)\tau_{2,3}$, so in fact the syzygies are generated by $\tau_{2,3}$. (In this simple case it is easy to see directly that $\tau_{2,3}$ generates the syzygies on g_1, g_2: Just use unique factorization and the fact that g_1 and g_2 are relatively prime.)

One corollary of Theorem 15.10 is a sharpened form of the Hilbert syzygy theorem, which says that every finitely generated S-module has a free resolution of length $\leq r$. We will give a more abstract proof in Chapter 19.

Corollary 15.11. *With notation as in Theorem* 15.10 *suppose that the* g_i *are arranged so that whenever* $\text{in}(g_i), \text{in}(g_j)$ *involve the same basis vector* e *of* F, *say* $\text{in}(g_i) = n_i e$ *and* $\text{in}(g_j) = n_j e$ *with* $n_i, n_j \in S$, *we have*

$$i < j \Rightarrow n_i > n_j \quad \textit{in the lexicographic order.}$$

If the variables $x_1, \ldots, x_s$ *are missing from the initial terms of the* g_i, *then the variables* $x_1, \ldots, x_{s+1}$ *are missing from the* $\text{in}(\tau_{ij})$, *and* $F/(g_1, \ldots, g_t)$ *has a free resolution of length* $\leq r - s$.

In particular, every finitely generated S*-module has a free resolution of length* $\leq r$.

Remark: The last statement is true for all S-modules: The Hilbert syzygy theorem with Auslander's lemma (Theorem A3.18) shows that every S-module has a projective resolution of length $\leq r$, and every projective S-module is free. (See, for example, Lam [1978] for an exposition.)

Proof. By Theorem 15.10 we have $(\tau_{ij}) = m_{ji}\varepsilon_i$, where $m_{ji} = m_j/\text{GCD}(m_i, m_j)$. Since $(m_i = \text{in}(g_i)) \geq (m_j = \text{in}(g_j))$ in the lexicographic order, x_{s+1} appears with at least as high a power in m_i as in m_j and thus does not appear at all in m_{ji}. This proves the first statement.

We next show that $F/(g_1, \ldots, g_t)$ has a free resolution of length $\leq r - s$ by induction on $r - s$. Suppose first that $r - s = 0$, so that none of the

variables x_i appears in the terms $\mathrm{in}(g_i)$; we must show that $F/(g_1, \ldots, g_t)$ is free.

Since the initial terms of the g_i must be scalars times basis elements of F, we see that $\mathrm{in}(g_1, \ldots, g_t)$ is the free submodule of F generated by the e_i that appear among the $\mathrm{in}(g_i)$. Let F' be the free submodule spanned by the other e_j, and consider the composite map

$$F' \subset F \to F/(g_1, \ldots, g_t).$$

By Theorem 15.3, $F/(g_1, \ldots, g_t)$ has a basis consisting of precisely the monomials coming from F', so the map is an isomorphism and $F/(g_1, \ldots, g_t) \cong F'$ is free as required.

Now suppose $r - s > 0$. By the first statement, $x_1, \ldots, x_{s+1}$ are missing from the initial terms of the τ_{ij}. We may order the τ_{ij} so as to satisfy the same hypothesis as that on the g_i. It follows from the induction that $\oplus S\varepsilon_i/(\{\tau_{ij}\})$ has a free resolution of length $\leq r-s-1$. Putting this together with the map $\oplus S\varepsilon_i \to F$, we get the desired free resolution of $F/(g_1, \ldots, g_t)$. □

15.6 History of Gröbner Bases

The earliest use of what amounts to the existence of Gröbner bases may be that of Gordan [1900, pp. 141–156]. Gordan uses Gröbner bases ("le système irreducible N" on page 152 is one) and the finite generation of monomial ideals to deduce Hilbert's basis theorem, just as in Exercise 15.15.

A major step toward the theory presented in this chapter was taken by Macaulay, who introduced total orderings of the set of monomials of a ring [1927] and used them to characterize the possible Hilbert functions of graded ideals by comparing them with monomial ideals.

Gröbner published applications of Macaulay's idea of ordering monomials and explicitly finding a basis for a zero-dimensional factor ring as early as [1939], though his use of them apparently goes back even earlier, perhaps to 1932. In a passage from a paper on elimination theory [1950], he wrote, "I have used and tested these methods for about 17 years in the most varied and complicated cases, and I believe that I can say on the basis of my experience that they represent in all cases a useful and worthwhile tool for the solution of these and similar ideal-theoretic problems." In 1964 he proposed that his student, Bruno Buchberger, compute such bases as a thesis problem. As seems to have been his practice in some other cases[1] as

[1]Wolfgang Vogel was a postdoctoral student of Buchberger in Innsbruck at about the time Buchberger did his work. He tells of another of Gröbner's students, whose thesis problem required computing a certain free resolution. Vogel was interested in such computations, and discussed the problem with the student. Later, at dinner

well, he did not mention to Buchberger that he already had a solution to the problem! It was not until 1984 that Buchberger learned the early part of the story (see Buchberger [1987] for this and related history).

As Gröbner must have hoped, Buchberger's solution to his thesis problem contained ideas going well beyond what Gröbner himself had known. The thesis [1965; University of Innsbruck] contains Buchberger's criterion and algorithm (our Theorem 15.8 and Algorithm 15.9) in implicit form. The essential added ingredient was the notion of critical pairs. Buchberger made his ideas more explicit and usable in his [1970] and [1976] papers.

There were several independent streams of activity that produced similar methods and algorithms. Hironaka used a division algorithm closely related to the one we have presented in his landmark paper on resolution of singularities [1964]. He introduced "standard bases," which are analogous to what we have called Gröbner bases, following a now more common usage. It is worth noting that Hironaka's work was done for power series (with questions of convergence treated), which in some ways is a deeper form of the division algorithm than the one treated here. He thought of it as generalizing the classical Weierstrass preparation and division theorems for convergent power series in one variable.

Grauert [1972] independently introduced standard bases and a division algorithm in power series rings, applying them to the construction of versal deformation spaces. Grauert also examines in this paper the effect of a general change of coordinates.

Bergman studied a more general version of Gröbner bases, aimed at associative (noncommutative) algebras and still more general algebraic systems [1978, especially Section 10.3]. Bergman's ideas specialize to Buchberger's algorithm in the commutative case. He remarked that the ideas had already been used—and called "obvious"—by Cohn [1966] and others. Other sources for the noncommutative theory include Priddy [1970] and Knuth-Bendix [1967].

Spear [1977] and Schreyer [1980] seem to be the first to have written down a method for the computation of syzygies by means of the division algorithm. (Spear's work, written as a report on a package he was developing for Macsyma, contains no mathematical details.) The formulation of Theorem 15.10 and the proof of the Hilbert syzygy theorem that we have given are Schreyer's.

15.7 A Property of Reverse Lexicographic Order

The reverse lexicographic order on S satisfies a key property not shared by other orders that makes the connection between an ideal and its initial

---

with Gröbner, Vogel remarked that the computation was quite difficult. "I know," replied Gröbner. "I need confirmation of the result!"

ideal particularly tight. As Bayer and Stillman show [1987a], it also has practical consequences that make the reverse lexicographic order preferable for computation in some circumstances.

Since we wish to be able to work with modules, we need the following definition.

Definition. *Let F be a graded free S-module with basis $\{e_1, \ldots, e_n\}$. A monomial order $>$ on F is called a* ***reverse lexicographic order*** *if it refines the order by total degree and satisfies the following property: If $f \in F$ is a homogeneous element and* $\operatorname{in}(f) \in (x_s, \ldots, x_r)F$ *for some $1 \leq s \leq r$, then $f \in (x_s, \ldots, x_r)F$.*

Equivalently, as the reader may check, a reverse lexicographic order is defined by choosing an order on the e_i, say $e_1 > \ldots > e_n$, and setting $me_i > ne_j$ iff either $\deg me_i > \deg ne_j$ or the degrees are the same and $m >_{\text{revlex}} n$ or $m = n$ and $i < j$.

The defining property of reverse lexicographic orders translates into good behavior upon factoring out the last variable. The following easy result is the key.

Proposition 15.12. *Suppose that F is a free S-module with basis $\{e_1, \ldots, e_n\}$ and reverse lexicographic order, and suppose that $g_1, \ldots, g_t$ is a homogeneous Gröbner basis of a graded submodule M.*

a. $\operatorname{in}(M + x_rF) = \operatorname{in}(M) + x_rF$. *Thus $g_1, \ldots, g_t, x_re_1, \ldots, x_re_n$ is a Gröbner basis of $M + x_rF$.*

b. $(\operatorname{in}(M) :_F x_r) = \operatorname{in}(M :_F x_r)$. *Further, if we set*

$$\tilde{g}_i = g_i/(GCD(x_r, g_i)),$$

then $\tilde{g}_1, \ldots, \tilde{g}_t$ is a Gröbner basis for $(M :_F x_r)$.

The proposition remains true, by virtually the same proof, if x_r is replaced by x_r^d. See Exercise 15.41 for an application.

Proof.

a. It suffices to show that $\operatorname{in}(M+x_rF) \subset \operatorname{in}(M)+x_rF$, the other inequality being clear. Suppose $f = g + x_rh$ with $g \in M$ and $h \in F$. We must show that $\operatorname{in}(f) \in \operatorname{in}(M) + x_rF$. If $\operatorname{in}(f)$ is divisible by x_r, then we are done. Otherwise, $\operatorname{in}(f)$ must be one of the terms of g, and this term is greater than any term of x_rh. Thus $\operatorname{in}(f) = \operatorname{in}(g) \in \operatorname{in}(M)$.

b. If x_r divides $\operatorname{in}(g)$ for some homogeneous $g \in F$, then since we are using reverse lexicographic order, x_r divides g. The first statement of b follows at once.

By the same reasoning, $(\text{in}(g_i) :_F x_r)$ is generated by $\text{in}(\tilde{g}_i)$ for every i, whence

$$(\text{in}(M) :_F x_r) = (\text{in}(\tilde{g}_1), \ldots, \text{in}(\tilde{g}_t)).$$

Since clearly $\tilde{g}_1, \ldots, \tilde{g}_t \in (M :_F x_r)$, this shows that the $\text{in}(\tilde{g}_i)$ form a Gröbner basis. □

Using these properties, we get a criterion for x_r to be a nonzerodivisor on an S-module, or more generally for the last variables in reverse order to be a regular sequence. (Recall from Chapter 10 that $x_r, \ldots, x_s$ form a regular sequence on an S-module N if, first, $(x_r, \ldots, x_s)N \neq N$ and, second, x_r is a nonzerodivisor on N, x_{r-1} is a nonzerodivisor on N/x_rN, and so on.)

Theorem 15.13 (Bayer and Stillman [1987]). *Let F be a free module with basis and a reverse lexicographic monomial order. Suppose $M \subset F$ is a graded submodule. The elements $x_r, \ldots, x_s$ form a regular sequence on F/M iff $x_r, \ldots, x_s$ form a regular sequence on $F/\text{in}(M)$.*

These results may be used to show that certain homological properties of F/M may be deduced from $F/\text{in}(M)$; see Corollary 19.11 and Corollary 20.21.

We note that if M is a graded submodule of F then any permutation of a regular sequence of F/M is again a regular sequence on F/M. Thus we could make the same statement with the variables in the natural order $x_s, \ldots, x_r$. But this "permutability of regular sequences" is somewhat subtle: It is not true without either local or graded hypotheses. We shall return to this issue in Chapter 17.

Before proving the theorem we need the following elementary criterion.

Proposition 15.14. *Let F be a free module with basis $\{e_1, \ldots, e_n\}$. If $N \subset F$ is a monomial submodule minimally generated by $n_1, \ldots, n_t$, then a sequence of monomials $m_1, \ldots, m_u \in S$ is a regular sequence modulo N iff each m_i is relatively prime to each n_l and to each m_j for $j \neq i$.*

Proof. Suppose first that each m_j is relatively prime to each n_l and to each m_i for $j \neq i$. Since all the m_i and n_l are monomials, any polynomial annihilating m_v modulo $N + (m_1, \ldots, m_{v-1})F$ is a sum of monomials from the sets $(Sn_l : m_v) = Sn_i$ and $(m_iF : m_v) = m_iF$. This shows that $m_1, \ldots, m_u$ is a regular sequence on F/N.

Conversely, suppose that $m_1, \ldots, m_u$ is a regular sequence on F/N. We will do induction on u. First we show that m_1 is relatively prime to each n_l. If $\text{GCD}(m_1, n_l) = n$, then $m_1 n_l/n \in N$, and since m_1 is a nonzerodivisor on F/N, we see that $n_l/n \in N$. Since n_l is part of a minimal set of generators of N, we must have $n_l/n = n_l$, so $n = 1$.

Since, in addition, no $m_1 e_l$ is in N, it is immediate that $n_1, \ldots, n_t$, $m_1e_1, \ldots, m_1e_n$ is a minimal set of generators for $N + m_1F$. Now $m_2, \ldots, m_u$ satisfy the hypothesis of the proposition with respect to the submodule

$N + m_1 F$, so by induction these m_i are relatively prime to each other, to each n_l, and to each $m_1 e_i$. From the last condition we deduce that they are relatively prime to m_1, and we are done. □

The next result is a generalization of one implication of Theorem 15.13. We will say that a monomial order $>$ on a free S-module F with basis $\{e_i\}$ is compatible with a monomial order $>$ on S itself if for $h \in S$ and $f \in F$ we have $\operatorname{in}(hf) = \operatorname{in}(h)\operatorname{in}(f)$. Equivalently, a compatible monomial order on F is one that compares monomials me_i and $m'e_i$ involving the same basis vector by using the given ordering on m and m'. Most monomial orders used in practice have this property.

Proposition 15.15. *Let F be a free S-module with basis and monomial order compatible with a given monomial order on S. If $M \subset F$ is any submodule and $h_1, \ldots, h_u \in S$ are such that $\operatorname{in}(h_1), \ldots, \operatorname{in}(h_u)$ is a regular sequence on $F/\operatorname{in}(M)$, then $h_1, \ldots, h_u$ is a regular sequence on F/M, and $\operatorname{in}(M + (h_1, \ldots, h_u)F) = \operatorname{in}(M) + \sum_{i=1}^{u} \operatorname{in}(h_i)F$.*

Proof. By induction we may reduce at once to the case $u = 1$, and to simplify the notation we write h for h_1. We first show that h is a nonzerodivisor modulo M. Suppose $hf \in M$ for some $f \in F$. We must prove that $f \in M$, and we may do induction on the size of $\operatorname{in}(f)$. We have $\operatorname{in}(hf) = \operatorname{in}(h)\operatorname{in}(f) \in (M)$, so by our hypothesis $\operatorname{in}(f) \in \operatorname{in}(M)$. Thus $h(f - \operatorname{in}(f)) \in M$, and by our induction $f - \operatorname{in}(f) \in M$, so we are done.

Next we must show that if $g = hf + m$ with $m \in M$, then $\operatorname{in}(g) \in \operatorname{in}(M) + \operatorname{in}(h)F$. We do induction on the size of $\operatorname{in}(f)$. If $\operatorname{in}(hf) = \operatorname{in}(h)\operatorname{in}(f) \in \operatorname{in}(M)$, then since $\operatorname{in}(h)$ is a nonzerodivisor modulo $\operatorname{in}(M)$ we have $\operatorname{in}(f) = \operatorname{in}(m')$ for some m' in M. We may replace the given expression for g by the expression $g = h(f - m') + (m + hm')$. By induction we see that $\operatorname{in}(g) \in \operatorname{in}(M) + \operatorname{in}(h)F$.

Thus we may suppose that $\operatorname{in}(h)\operatorname{in}(f) \notin \operatorname{in}(M)$. In particular, the terms $\operatorname{in}(hf) = \operatorname{in}(h)\operatorname{in}(f)$ and $\operatorname{in}(m)$ involve distinct monomials. Whichever of these is greater cannot cancel against any other term in $hf + m$, and thus is the initial term of $hf + m$, so we are done. □

Proof of Theorem 15.13. If $x_r, \ldots, x_s$ is a regular sequence on $F/\operatorname{in}(M)$, then $x_r, \ldots, x_s$ is a regular sequence on F/M by Proposition 15.15.

It remains to prove the converse. In the case where $s = r$, Proposition 15.12b shows that x_r is a nonzerodivisor on $F/\operatorname{in}(M)$, as required. Factoring out $x_r F$ and using Proposition 15.12a, we are done by induction. □

15.8 Gröbner Bases and Flat Families

All the applications of the idea of Gröbner bases work by comparing an arbitrary ideal with its "initial" ideal, which is a monomial ideal (and more generally, by comparing a submodule of a free module with an associated initial submodule). Why should these two be similar enough to make the comparison profitable? The situation is quite similar to that of the associated graded ring treated in Chapter 5. As in the case of the Rees algebra construction defined in Chapter 6, an "explanation" is provided by the existence of certain flat families, which we will now describe. For simplicity we give the constructions below for ideals, rather than for arbitrary submodules of a free module with basis; the (easy) extension to the case of modules is left to the interested reader.

The flat family that we will describe is defined in terms of an integral weight function $\lambda : \mathbf{Z}^r \to \mathbf{Z}$. For convenience of notation, we think of λ as a function on monomials, and if $m = x^a$, we write $\lambda(m) \in \mathbf{Z}$ in place of $\lambda(a)$. Let $>_\lambda$ be the weight order defined by λ on the monomials of S. Although it is only a partial order, much of our formalism for monomial orders can be imitated for $>_\lambda$. For example, given $g \in S$ we write $\mathrm{in}_\lambda(g)$ for the sum of all the terms of g that are maximal with respect to $>_\lambda$, and if I is an ideal we write $\mathrm{in}_\lambda(I)$ for the ideal generated by $\mathrm{in}_\lambda(g)$ for all $g \in I$.

Before describing the flat family, we will show that integral weight orders are potent enough to capture the transition from a given ideal to its initial ideal with respect to an arbitrary monomial order. Suppose that $>$ is a monomial order on S, and $I \subset S$ is an ideal. Given any finite set of pairs of monomials $\mathcal{S} = \{(m_i > n_i)\}$, Exercise 15.12 shows that there is an integral weight order $>_\lambda$ such that $m_i >_\lambda n_i$ for all i. Thus we may apply the following proposition.

Proposition 15.16. *Let $>$ be a monomial order on S, and suppose that $g_1, \ldots, g_t$ is a Gröbner basis for an ideal I with respect to $>$. There is a finite set $\mathcal{S} = \{(m_1 > n_1), \ldots, (m_s > n_s)\}$ of pairs of monomials such that if $>_\lambda$ is a weight order on S with $(m_1 >_\lambda n_1), \ldots, (m_s >_\lambda n_s)$, then $g_1, \ldots, g_t$ is a Gröbner basis for I with respect to $>_\lambda$ and $\mathrm{in}_\lambda(I) = \mathrm{in}_>(I)$.*

Proof. For each $i = 1, \ldots, t$ we put into $\mathcal{S}$ all the pairs of monomials of the form $(\mathrm{in}(g_i) > n)$, where n is a noninitial monomial of g_i. (Here for simplicity we abuse our notation by writing $\mathrm{in}(g_i)$ for the initial monomial of g_i instead of the initial term.) Next, we use the Buchberger criterion, Theorem 15.8, to verify that the g_i are a Gröbner basis with respect to $>$; the verification depends on computing the initial terms of finitely many polynomials and involves finitely many uses of the division algorithm. For each polynomial g whose initial term we must compute, we put the pairs $(\mathrm{in}_>(g), n)$ into $\mathcal{S}$ for every noninitial monomial n of g. Similarly, each use of the division algorithm involves finitely many comparisons of pairs

of monomials. We expand the list of monomials by including the pairs of monomials involved.

Now the division algorithm and the proof of Buchberger's criterion work for weight orders and other monomial partial orders satisfying the multiplicative properties in the definition of monomial orders just as well as for total orders, so long as all the initial terms involved are monomials. Thus a second use of the Buchberger algorithm shows that the g_i form a Gröbner basis with respect to $>_\lambda$. In particular, the $\mathrm{in}(g_i)$ generate $\mathrm{in}_\lambda(I)$. Since $\mathrm{in}_>(g_i) = \mathrm{in}_\lambda(g_i)$, we are done. □

We may describe the flat family of algebras informally as follows. Let λ be an integral weight function. For any $0 \neq t \in k$, there is an automorphism of S carrying x_i to $t^{-\lambda(x_i)}x_i$, and we write I_t for the image of I under this automorphism. Clearly, all the rings S/I_t for $t \neq 0$ are isomorphic. But as t approaches 0, the initial terms of polynomials in I_t—those whose values under λ are largest—come to dominate the polynomials, and the limit, the fiber over $t = 0$, will be $S/\mathrm{in}_\lambda(I)$.

To make precise mathematics out of this description, let $S[t]$ be a polynomial ring in one variable over S. For any $g \in S$, we define $\tilde{g} \in S[t]$ as follows. Write $g = \sum u_i m_i$, where the m_i are monomials and $0 \neq u_i \in k$. Let $b = \max \lambda(m_i)$, and set

$$\tilde{g} = t^b g(t^{-\lambda(x_1)}x_1, \ldots, t^{-\lambda(x_r)}x_r).$$

Because of the way b was defined, we see that $\tilde{g}$ is $\mathrm{in}_\lambda(g)$ plus t times a polynomial in t and $x_1, \ldots, x_r$. For any ideal $I \subset S$, let $\tilde{I}$ be the ideal of $S[t]$ generated by $\{\tilde{g} | g \in I\}$. It follows that $S[t]/((t) + \tilde{I}) \cong S/\mathrm{in}_\lambda(I)$. The next result extends this and gives a more sophisticated interpretation.

Theorem 15.17. *For any ideal $I \subset S$, the $k[t]$-algebra $S[t]/\tilde{I}$ is free—and thus flat—as a $k[t]$-module. Furthermore,*

$$S[t]/\tilde{I} \otimes_{k[t]} k[t, t^{-1}] \cong S/I[t, t^{-1}],$$

while

$$S[t]/\tilde{I} \otimes_{k[t]} k[t]/(t) \cong S/\mathrm{in}_\lambda(I).$$

Thus $S[t]/\tilde{I}$ is a flat family over $k[t]$ of quotients of S whose fiber over 0 is $S/\mathrm{in}_\lambda(I)$ and whose fiber over any $(t-u)$, for $0 \neq u \in k$, is S/I.

(To give the flat family constructively, in the spirit of this chapter, we should specify a finite set of generators for the ideal $\tilde{I}$. Results of this sort may be found in Exercise 15.25.)

Proof. From the fact that $\tilde{g}$ is $\mathrm{in}_\lambda(g)$ plus t times a polynomial in t and $x_1, \ldots, x_r$, it is clear that

$$S[t]/\tilde{I} \otimes_{k[t]} k[t]/(t) = S[t]/(\tilde{I} + (t)) = S/\mathrm{in}_\lambda(I).$$

Let φ be the automorphism of $S \otimes_{k[t]} k[t, t^{-1}] = S[t, t^{-1}]$ defined by $\varphi(x_i) = t^{\lambda(x_i)} x_i$. This automorphism takes the ideal $\tilde{I}S[t, t^{-1}]$ to the ideal $IS[t, t^{-1}]$; it follows that φ induces an isomorphism $S[t]/\tilde{I} \otimes_{k[t]} k[t, t^{-1}] \cong S/I[t, t^{-1}]$.

It thus remains to prove the first statement of the theorem. Let $>$ be a monomial order refining $>_\lambda$, and let B be the set of monomials not $\text{in}_>(I)$. B is a basis of S/I by Theorem 15.3, and we claim that B is also a $k[t]$-basis for $S[t]/\tilde{I}$.

First, to prove linear independence, it is enough to show that the elements of B are linearly independent over $k[t, t^{-1}]$, as elements of $S[t, t^{-1}]$. From Theorem 15.3 we deduce that the elements of B form a $k[t, t^{-1}]$-basis of $S[t, t^{-1}]/IS[t, t^{-1}]$. Thus $\varphi^{-1}(B)$ is a basis for $S[t, t^{-1}]/\tilde{I}S[t, t^{-1}]$. But the automorphism φ^{-1} carries any monomial m into $t^{-\lambda(m)}m$, that is, a unit of $S[t, t^{-1}]$ times m. Thus B itself is a $k[t, t^{-1}]$-basis of $S[t, t^{-1}]/\tilde{I}S[t, t^{-1}]$. In particular, its elements are linearly independent in $S[t]/\tilde{I}$.

Finally, we must show that B generates $S[t]/\tilde{I}$ as a $k[t]$-module.

Regarding B as a subset of $S[t]$, we must show that the $k[t]$-span of B contains, modulo elements of $\tilde{I}$, every monomial m in the x_i. Because the order $>$ is Artinian, we may inductively assume that this has been verified for every monomial $n < m$. The monomial m is either in B or else $m = \text{in}_>(g)$ for some $g \in I$. In the latter case $m - \tilde{g}$ is a $k[t]$-linear combination of monomials that are $< m$, and we are done by induction. □

The technique embodied in the preceding result can be used to give a flat family connecting any given finite set of ideals to their initial ideals. (If one is willing to exchange the simple "base" $k[t]$ of the family used above for something more complicated—generally non-Noetherian—one can do it for all ideals at once—see Exercise 15.26.)

We shall now give some pictorial examples. We treat the case of three points in the projective plane (Figures 15.3 and 15.4) and the case of a smooth conic in the projective plane (Figures 15.5 and 15.6) first with the lexicographic and then the reverse lexicographic orders. We thus work in the polynomial ring in three variables, $k[x, y, z]$, with $x > y > z$. The coordinate triangle of lines $x = 0, y = 0$, and $z = 0$ is distinguished by the choice of coordinates, shown in Figure 15.2.

To simulate the lexicographic order we use two weight orders. First we deform according to the family corresponding to weights (1, 0, 0). This may be interpreted either as "attract to $x = 0$" or as "repel from $y = z = 0$." Next we deform according to the family corresponding to weights (0, 1, 0). This may be interpreted either as "attract to $y = 0$" or as "repel from $x = z = 0$."

Similarly, to simulate the reverse lexicographic order we use first the deformation corresponding to the weight vector (1, 1, 0), or equivalently $(0, 0, -1)$. Its effect is to attract to the point $x = y = 0$ and to repel from the line $z = 0$. Next we use the weight vector (1, 0, 1) or equivalently $(0, -1, 0)$. The effect is to attract to $x = z = 0$ and to repel from $y = 0$.

When looking at the figures, bear in mind that in each deformation each corner of the coordinate triangle is fixed under the deformations, and each of the three lines of the triangle is sent into itself.

Here is the case of a set Γ of 3 general points in the plane. If we take these to be the points (1, 1, 1), (1/3, 1/2, 1), and (1/2, 1/3, 1), then the ideal of Γ is

$$\begin{aligned} I(\Gamma) = (&x^2 + xy - (11/6)xz - yz + (5/6)z^2, \\ &xy + y^2 - xz - (11/6)yz + (5/6)z^2, \\ &y^2 - (2/7)xz - (47/42)yz + (17/42)z^2). \end{aligned}$$

With a little computation one sees that in lexicographic order the initial ideal is $(x^2, xy, xz, y^3) = (x(x, y, z), y^3)$. If we remove the primary component at the irrelevant ideal we get (x, y^3). Thus the limiting position for this deformation is the second-order neighborhood of the point $x = y = 0$ in the

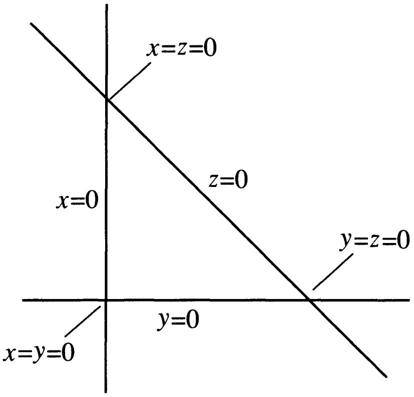


FIGURE 15.2.

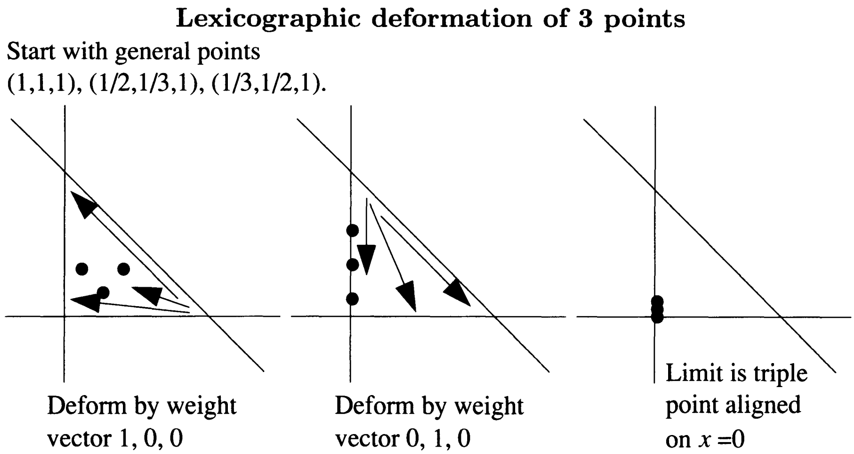


FIGURE 15.3.

Reverse lexicographic deformation of 3 points

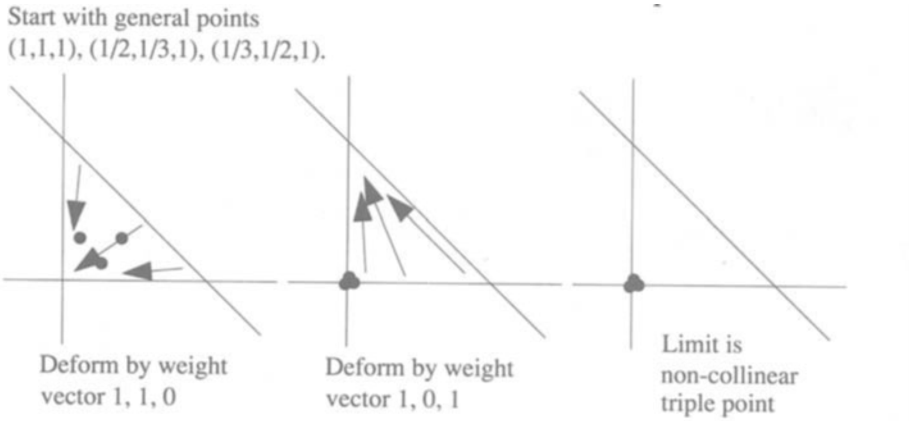


FIGURE 15.4.

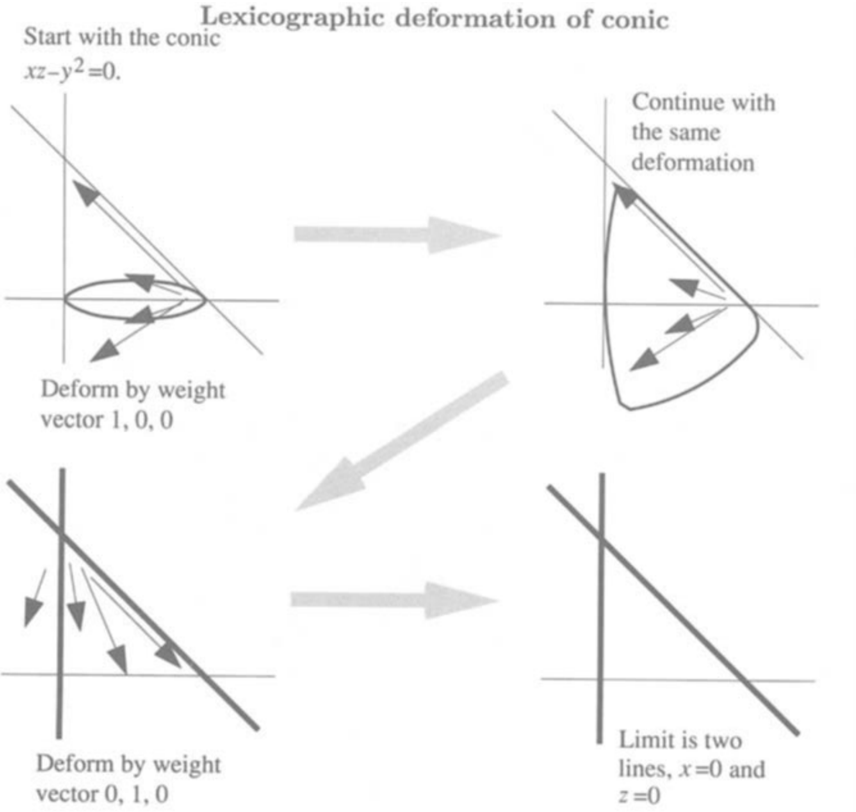


FIGURE 15.5.

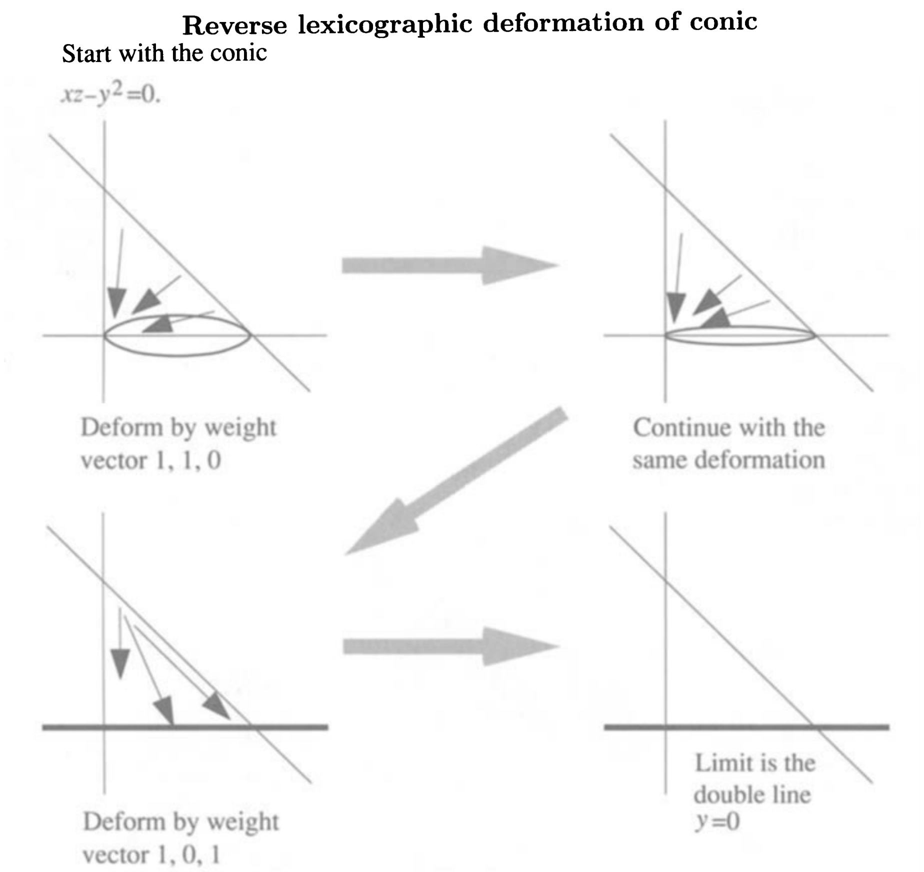


FIGURE 15.6.

line $x = 0$. It is perhaps easier to see that in reverse lexicographic order the initial ideal is (x^2, xy, y^2), so the limiting position is the first-order neighborhood of the point $x = y = 0$ in the plane. From these computations we see that the two deformations have nonisomorphic limits.

Next we try the same deformations on the conic with equation $xz - y^2 = 0$. We draw this as an ellipse tangent to the lines $x = 0$ and $z = 0$ along the line $y = 0$. In lexicographic order the initial term is xz, corresponding to a limiting position that is the union of the lines $x = 0$ and $z = 0$. In reverse lexicographic order, on the other hand, the initial term of the equation is y^2, corresponding to the double line with reduced line $y = 0$. We have added a picture of the stage in which the first deformation is merely approaching its limit.

There is more to be seen in Figures 15.5 and 15.6. For example, as a conic undergoes the degeneration corresponding to one of the two orders, the dual conic (the set of its tangent lines) undergoes the other. This phenomenon is "visible" in the pictures; can the reader spot it?

15.9 Generic Initial Ideals

So far we have always considered Gröbner bases with respect to both some fixed set of variables in a polynomial ring and a fixed set of generators of a free module. The results of Gröbner basis computations depend heavily on the choice of variables and basis made. By allowing a generic change of basis and coordinates, we may eliminate this dependence and get a **generic initial ideal** that depends only on a choice of monomial order. Some of the properties of generic initial ideals were exploited by Hartshorne [1966] to prove the connectedness of Hilbert schemes. We will prove stronger properties below; these sections should give the reader a good preparation for the algebraic part of Hartshorne's paper. Generic initial ideals were also considered by Grauert [1972] in the case of power series rings. He seems to have been the first to observe that the generic initial ideal is a combinatorial invariant that contains quite a lot of information. Generic initial ideals have also been exploited to bound the invariants of projective varieties (see Cook [in press] and Braun and Fløystad [in press]).

To get a sense of the information contained in the generic initial ideal, suppose $I \subset S$ is an ideal and we take reverse lexicographic order on S. In generic coordinates we can read off from $\text{in}(I)$ the depth of S/I (= the largest t such that $x_{r-t+1}, \ldots, x_r \notin \text{in}(I)$) and the regularity of I (= the regularity of $\text{in}(I)$; in characteristic 0 this is just the maximal degree of a minimal generator of $\text{in}(I)$), as well as things like the Hilbert function of S/I that we could read off from $\text{in}(I)$ in any coordinate system. See Chapters 18 and 20 and Bayer and Stillman [1987b] for more information.

In this section we will explain the basic facts about generic initial ideals. The first treatment, in characteristic 0, is that of Galligo [1974]. Bayer and Stillman [1987a] did the theory in arbitrary characteristic. The combinatorial analysis and the properties of Borel-fixed ideals in characteristic p were worked out by Pardue [1994], who has given a treatment covering many other group actions. We shall follow his treatment here, adapted to our special case. Although everything we do can be extended to the case of submodules of a graded free module with basis, for simplicity we will stick to the case of ideals.

Throughout this section we will work with a fixed monomial order $>$ on $S = k[x_1, \ldots, x_r]$ that refines the partial order by degree and that satisfies $x_1 > \ldots > x_r$. We assume that the ground field k is infinite. All ideals considered will be homogeneous.

It is convenient to speak of taking initial ideals with respect to a given coordinate system and order, so instead of making a generic transformation of coordinates, we will transform an ideal by a generic linear transformation and take its initial ideal in the given coordinates.

We begin by establishing some notation for the groups of transformations that we will use. The general linear group $\mathcal{G} := \text{GL}(r, k)$ of invertible $r \times r$ matrices over k acts as a group of algebra automorphisms on S by acting

on the variables: If g is the matrix with (i,j) entry g_{ij}, we define $g(x_j)$ to be $\sum_i g_{ij}x_i$. If $m = \prod_j x_j^{a_j}$ is an arbitrary monomial of S then $g(m)$, which we shall often write as gm, is given by $g(m) = \prod_j(\sum_i g_{ij}x_i)^{a_j}$.

Because we have distinguished an ordering of the variables, certain subgroups of $\mathcal{G}$ play an important role. Let $\mathcal{B}$ be the **Borel subgroup** of $\mathcal{G}$ consisting of upper triangular invertible matrices, and let $\mathcal{B}'$ be the group of invertible lower triangular matrices. Let $\mathcal{U} \subset \mathcal{B}$ be the **unipotent subgroup** consisting of upper triangular matrices with ones on the diagonal. $\mathcal{U}$ is generated by the **elementary upper triangular matrices** γ_{ij}^c for $i < j$ and $c \in k$, where $\gamma_{ij}^c x_j = cx_i + x_j$ and $\gamma_{ij}^c x_u = x_u$ for $u \neq j$. Similarly, $\mathcal{B}'$ is generated by the diagonal matrices and the **elementary lower triangular matrices** $\gamma_{ij}'^c$ for $1 \leq i < j \leq r$ whose action is given by $\gamma_{ij}'^c x_i = x_i + cx_j$ and $\gamma_{ij}'^c x_u = x_u$ for $u \neq i$.

If $V \subset S_d$ is a t-dimensional space of forms of degree d, then we may represent V as a one-dimensional subspace $L = \wedge^t V \subset \wedge^t S_d$: If V has basis $f_1, \ldots, f_t$, then the subspace L is spanned by $f := f_1 \wedge \ldots \wedge f_t$. The reader unfamiliar with multilinear algebra will find more information in Appendix 2. We define a **monomial** of $\wedge^t S_d$ to be an element of the form $n = n_1 \wedge \ldots \wedge n_t$, where the n_i are degree-d monomials of S. If the n_i are not distinct, then $n = 0$; in the contrary case the line kn determines and is determined by the finite set $\{n_1, \ldots, n_t\}$. We define a **term** in $\wedge^t S_d$ to be a product $a \cdot n$, where $a \in k$ and n is a monomial. We will say that $a \cdot n = a \cdot n_1 \wedge \ldots \wedge n_t$ is a **normal expression** if the n_i are ordered so that $n_1 > \ldots > n_t$.

We order the monomials of $\wedge^t S_d$ by ordering their normal expressions lexicographically. That is, if $n = n_1 \wedge \ldots \wedge n_t$ and $n' = n'_1 \wedge \ldots \wedge n'_t$ are normal expressions, then $n > n'$ iff $n_i > n'_i$ for the smallest i such that $n_i \neq n'_i$. As usual, we extend the order to terms, and define the initial term of an element $f \in \wedge^t S_d$ to be the greatest term with respect to the given order.

Write m_i for $\mathrm{in}(f_i)$. We may replace the f_i by some linear combinations of themselves (without changing V) to ensure that the m_i are distinct and that $m_1 > \ldots > m_t$. With this choice, $m_1 \wedge \ldots \wedge m_t$ is the normal expression for the initial term of f.

15.9.1 Existence of the Generic Initial Ideal

Theorem 15.18. *Let $I \subset S$ be a homogeneous ideal. There is a Zariski open set $U = \mathcal{B}'\mathcal{U} \subset \mathcal{G}$, meeting $\mathcal{U}$ nontrivially, and a monomial ideal $J \subset S$ such that for all $g \in U$ we have $\mathrm{in}(gI) = J$. For each $d \geq 0$, if the degree-d part J_d of J has dimension t, then $\wedge^t J_d$ is spanned by the greatest monomial of $\wedge^t S_d$ that appears in any $\wedge^t(gI_d)$ with $g \in \mathcal{G}$.*

Definition. *With I, J as in Theorem 15.18, J is called the* ***generic initial ideal*** *of I, written $J = \mathrm{Gin}(I)$.*

The significance of the assertion $U = \mathcal{B}'U$ is that all the action takes place in the coset space $\mathcal{B}'\backslash\mathcal{G}$. This is a much-studied object, which may be identified with the space of complete flags of linear subspaces of $\mathbf{A}^r$.

Proof. First consider the degree-d part I_d of I. Let $f_1, \ldots, f_t$ be a basis for I_d. If $h = (h_{ij})$ is a matrix of indeterminates, then $h(f_1 \wedge \ldots \wedge f_t) = h(f_1) \wedge \ldots \wedge h(f_t)$ is a linear combination of monomials of $\wedge^t S_d$ with coefficients that are polynomials in the h_{ij}. Suppose that $m = m_1 \wedge \ldots \wedge m_t$ is the earliest monomial that appears with a nonzero coefficient, and let $p_d(h_{11}, \ldots, h_{rr})$ be that coefficient. Let U_d be the set of $g = (g_{ij}) \in \mathcal{G}$ such that $p_d(g_{11}, \ldots, g_{rr}) \neq 0$. The degree-$d$ part of the initial ideal of gI will be $(m_1, \ldots, m_t)$ iff $g \in U_d$. Write J_d for the subspace of S_d spanned by $m_1, \ldots, m_t$.

We next show that $J := \oplus J_d$ is an ideal. It suffices to show for each d that $S_1 J_d \subset J_{d+1}$. Since U_d and U_{d+1} are open and dense, there is an element $g \in U_d \cap U_{d+1}$. We have $\mathrm{in}(gI)_d = J_d$ and $\mathrm{in}(gI)_{d+1} = J_{d+1}$, and the assertion follows.

The ideal J satisfies the last statement of the theorem by definition, and we will show that $U = \cap_{d=1}^{\infty} U_d$ is Zariski open and dense in $\mathcal{G}$. Since each U_d is Zariski open and dense, it suffices to show that U is equal to a finite intersection of the U_d. Supposing that J is generated by forms of degree $\leq e$, we will show that in fact $U = \cap_{d=1}^{e} U_d$.

Suppose that $g \in \cap_{d=1}^{e} U_d$. We know that $\mathrm{in}(gI_d) = J_d$ for all $d \leq e$. Thus $\mathrm{in}(gI) \supset J$. Since $\dim_k J_d = \dim_k I_d = \dim_k (gI)_d$ for every d, we see that $\mathrm{in}(gI) = J$ as required.

We next show that $U = \mathcal{B}'U$. In fact, a little more is true:

Lemma 15.19. *If $I_d \subset S_d$ is a subspace of dimension t and $b \in \mathcal{B}'$, then* $\mathrm{in}(\wedge^t I_d) = \mathrm{in}(\wedge^t b I_d)$.

Proof. Since $\mathcal{B}'$ is generated by diagonal matrices and elementary lower triangular matrices, it suffices to check the assertion when b is of one of these types. Choose a basis $f_1, \ldots, f_t$ for I_d and let $m_i = \mathrm{in}\, f_i$. Changing basis if necessary, we may assume that $m_1 > \ldots > m_t$. The diagonal matrices simply alter the coefficients of the terms of $f = f_1 \wedge \ldots \wedge f_t$ by nonzero scalars, so the assertion is true if b is diagonal.

Next suppose that $b = \gamma_{ij}^{\prime c}$ is an elementary lower triangular matrix. For any monomial $n = x_i^w m \in S_d$, where m is not divisible by x_i, bn is n plus a linear combination of monomials of the form $n' = x_i^{w-s} x_j^s m$ with $0 < s \leq w$. Since $x_i > x_j$, we see that each $n' < n$. Thus $\mathrm{in}(bf_i) = m_i$ for $1 \leq i \leq t$, so $\mathrm{in}(bf) = m_1 \wedge \ldots \wedge m_t = \mathrm{in}(f)$. □

To complete the proof of Theorem 15.18 we check that U meets the unipotent subgroup $\mathcal{U}$ nontrivially. The set $\mathcal{B}'\mathcal{U}$ is a dense open subset of $\mathcal{G}$; see Exercise 15.24 and its hint for a proof. Thus the dense set U contains

an element of the form bu with $b \in \mathcal{B}'$ and $u \in \mathcal{U}$. Since $U = \mathcal{B}'U$, it follows that $u = b^{-1}bu \in U$ as required. □

15.9.2 The Generic Initial Ideal is Borel-Fixed

The next result shows that generic initial ideals are quite special among monomial ideals. The description will be made explicit in Theorem 15.23.

Theorem 15.20 (Galligo, Bayer and Stillman). *If $I \subset S$ is a homogeneous ideal then* $\operatorname{Gin}(I)$ *is Borel-fixed in the sense that for all $g \in \mathcal{B}$, $g(\operatorname{Gin}(I)) = \operatorname{Gin}(I)$.*

Proof. Replacing I by gI for generic g, we may assume by Theorem 15.18 that $\operatorname{in}(I) = \operatorname{Gin}(I)$. Fix $i < j$, and let $\gamma_{ij}^1 = 1 + \gamma$ be an elementary upper triangular matrix, where γ is a strictly upper triangular matrix with a single nonzero entry. Along with diagonal matrices, such matrices generate the Borel group $\mathcal{B}$. Since the diagonal matrices stabilize any monomial ideal, it suffices to show that for each degree d we have $(1+\gamma)(\operatorname{in}(I_d)) = \operatorname{in}(I_d)$.

We may choose a basis $f_1, \dots, f_t$ for I_d with $\operatorname{in}(f_1) > \dots > \operatorname{in}(f_t)$. Let $f = f_1 \wedge \dots \wedge f_t$ be the corresponding generator of the one-dimensional subspace $\wedge^t I_d \subset \wedge^t S_d$. We have $\operatorname{in}(f) = \operatorname{in}(f_1) \wedge \dots \wedge \operatorname{in}(f_t)$.

If $(1+\gamma)(\operatorname{in}(I_d)) \neq \operatorname{in}(I_d)$ then $(1+\gamma)\operatorname{in}(f) \neq \operatorname{in}(f)$. Since γ is strictly upper triangular, the terms of $(1+\gamma)\operatorname{in}(f)$ other than $\operatorname{in}(f)$ are all strictly greater than $\operatorname{in}(f)$. Let am be one of these terms, where a is a nonzero scalar and m is a monomial of $\wedge^t S_d$. We shall show that for suitable diagonal matrices δ the monomial m appears with nonzero coefficient in $(1+\gamma)\delta f$. This will contradict the last statement of Theorem 15.18, proving that $(1+\gamma)(\operatorname{in}(I_d)) = \operatorname{in}(I_d)$ after all.

For each term $n = an_1 \wedge \dots \wedge n_t \in \wedge^t S_d$ we define the **weight** of n to be the monomial $w = \prod_s n_i \in S$. Let $f_w \in \wedge^t S_d$ be the sum of all the terms of f having weight w, so that we have $f = \sum_w f_w$. Let w_0 be the weight of $\operatorname{in}(f)$. Different terms of f may have the same weight, but $\operatorname{in}(f)$ is the unique term having weight w_0. If δ is a diagonal matrix and $\delta(x_i) = \delta_i x_i$ with $\delta_i \in k^\times$, then

$$\delta f = \sum_w w(\delta_1, \dots, \delta_r) f_w,$$

where $w(\delta_1, \dots, \delta_r) \in k^\times$ is the result of substituting δ_i for x_i in the monomial w. Thus

$$\begin{aligned}
(1+\gamma)\delta f &= \sum_w (1+\gamma)(w(\delta_1, \dots, \delta_r) f_w) \\
&= \sum_w w(\delta_1, \dots, \delta_r)(1+\delta) f_w \\
&= w_0(\delta_1, \dots, \delta_r)(1+\gamma)\operatorname{in}(f) + \sum_{w \neq w_0} w(\delta_1, \dots, \delta_r)(1+\gamma) f_w.
\end{aligned}$$

Thus the coefficient of m in $(1+\gamma)\delta f$ has the form

$$c(\delta_1,\ldots,\delta_r) := aw_0(\delta_1,\ldots,\delta_r) + \sum\nolimits_{w\neq w_0} a_w w(\delta_1,\ldots,\delta_r),$$

where the $a_w \in k$ is the coefficient of m in $(1+\gamma)f_w$. Since the term $aw_0(\delta_1,\ldots,\delta_r)$ is nonzero, we see that the polynomial c is nonzero. Since we have assumed that the ground field k is infinite, it follows that for sufficiently general values of $\delta_1,\ldots,\delta_n$, the value $c(\delta_1,\ldots,\delta_r)$ is nonzero, and this is what we had to prove. □

15.9.3 The Nature of Borel-Fixed Ideals

We next investigate the nature of Borel-fixed ideals. To treat the case of characteristic p it is useful to introduce a partial order $\prec_p$ on the natural numbers as follows: We say that $a \prec_p b$ if the binomial coefficient $\binom{b}{a} \not\equiv 0 (\operatorname{mod} p)$. Of course $\prec_0$ is the usual total order $\leq$. For $p > 0$, Gauss gave the following explicit description.

Proposition 15.21 (Gauss). *Suppose p is a prime number. We have $a \prec_p b$ iff each digit in the base-p expansion of a is $\leq$ the corresponding digit in the base-p expansion of b.*

The proof is immediate from a more refined result of Lucas.

Lemma 15.22 (Lucas). *If $a = \sum a_i p^i$ and $b = \sum b_i p^i$ with $0 \leq a_i, b_i < p$, then $\binom{b}{a} \equiv \Pi_i \binom{b_i}{a_i} (\operatorname{mod} p)$.*

Proof. Compare the coefficients of t^a in the expressions

$$\begin{aligned}(t+1)^b = (t+1)^{\Sigma b_i p^i} &= \prod (t+1)^{b_i p^i} \\ &\equiv \prod (t^{p^i}+1)^{b_i} (\operatorname{mod} p).\end{aligned}$$

□

We now give the combinatorial characterization of Borel-fixed ideals.

Theorem 15.23. *Let $J \subset S = k[x_1,\ldots,x_r]$ be an ideal, and let* $\operatorname{char} k = p \geq 0$.

a. *J is fixed by the group of diagonal matrices iff J is generated by monomials.*

b. *J is fixed by the group $\mathcal{B}$ of upper triangular matrices (that is, J is Borel-fixed) iff J is generated by monomials and the following condition is satisfied for all $i < j$ and all monomial generators m of J:*

   If m is divisible by x_j^t but by no higher power of x_j, then $(x_i/x_j)^s m \in J$ for all $i < j$ and $s \prec_p t$.

Proof.

a. Clearly any monomial ideal is fixed by the group of diagonal matrices. To prove the converse, suppose J is fixed by the diagonal matrices and let $f \in J$; it is enough to show that some monomial of f is in J. Choose a weight vector λ such that $\text{in}_\lambda(f)$ is a monomial; that is, only one monomial of f has maximal weight with respect to λ. We will show that $\text{in}_\lambda(f) \in J$.

   Let w be the weight of the term $\text{in}_\lambda(f)$. If we act on f with a diagonal matrix g_c having diagonal terms $(c^{-\lambda_1}, \ldots, c^{-\lambda_r})$, we replace each variable x_i by $c^{-\lambda_i}x_i$, so $\text{in}_\lambda(f)$ is multiplied by c^{-w}, and the other terms of f are multiplied by strictly less-negative powers of c. Thus we may write $c^w g_c f = \text{in}_\lambda(f) + cF(c,x)$ for some polynomial $F(c,x)$. Consider the morphism $\varphi : \mathbf{A}^1_k \to S$ defined by $\varphi(c) = c^w g_c f = \text{in}_\lambda(f) + cF(c,x)$. For $c \neq 0$ the matrix g_c is invertible. Since J is fixed under the group of diagonal matrices, $\varphi(c) \in J$ for $c \neq 0$. Since J is a Zariski closed subset, in fact a linear subspace, this implies that $\varphi(c) \in J$ for all c. (The fact that S is infinite-dimensional is not a problem: If J is the common zero locus of a set of linear functions $\alpha_i : S \to k$, then composing the α_i with φ we get polynomial functions from $\mathbf{A}^1_k$ to k that vanish simultaneously on precisely those c for which $\varphi(c) \in J$. Since these polynomials vanish for all nonzero c, they vanish for all c.)

b. If J is Borel-fixed then, by a, J is generated by monomials. If $m \in J$ is a monomial generator, we consider the action on m of an elementary upper triangular matrix $\gamma = \gamma^1_{ij}$. We may write $m = x_j^t m'$, where m' is not divisible by x_j, and we get

$$\gamma m = (cx_i + x_j)^t m' = \sum_{s \prec_p t} \binom{t}{s} (x_i/x_j)^s m.$$

   Since J is fixed under $\gamma = \gamma^c_{ij}$, we see that each $(x_i/x_j)^s m$ with $s \prec_p t$ is a monomial belonging to some polynomial in J. Being a monomial ideal, J contains all the monomials that appear in polynomials from J, and J thus contains the monomial $(x_i/x_j)^s m$ as required.

   Conversely, suppose J is a monomial ideal satisfying the condition in b. The formula above shows that for every monomial generator of J the polynomial γm is a sum of monomials in J. Since J is generated by monomials, $\gamma J = J$. Since the group of upper triangular matrices is generated by diagonal matrices and matrices γ^c_{ij}, we are done. □

A few examples will clarify the theorem. For simplicity we take only examples with all generators in a single degree. First, in characteristic 0: In two variables the Borel-fixed ideals are precisely the ideals generated by

"initial segments of the monomials" in each degree, such as $(x_1^3, x_1^2x_2, x_1x_2^2)$. But there are already more possibilities in three variables. For example, the ideals $(x_1^3, x_1^2x_2, x_1x_2^2)$ and $(x_1^3, x_1^2x_2, x_1^2x_3)$ are both Borel-fixed in any characteristic. In characteristic $p > 0$ any ideal of the form $(x_1^{p^e}, \dots, x_u^{p^e})$ is Borel-fixed. Products, intersections, sums, and quotients of Borel-fixed ideals are Borel-fixed, so it is easy to make further examples.

To exploit the results on generic initial ideals, we use the following fundamental property of Borel-fixed ideals.

Proposition 15.24 (Bayer and Stillman). *Suppose that $I \subset S = k[x_1, \dots, x_r]$ is a Borel-fixed ideal. For any $j = 1, \dots, r$ we have*

$$(I : x_j^\infty) = (I : (x_1, \dots, x_j)^\infty).$$

If char $k = 0$, *then in addition*

$$(I : x_j^s) = (I : (x_1, \dots, x_j)^s)$$

for every $s \geq 0$.

Proof. Suppose that for some integer s and some monomial m we have $x_j^s m \in I$. For the first statement it suffices to show that if $1 \leq i < j$ then for some $s' \geq s$ we have $x_i^{s'} m \in I$: For if s is sufficiently large, then

$$\begin{aligned}(I : x_j^\infty) &= (I : x_j^s)\\ &\subset (I : (x_1^{s'}, \dots, x_j^{s'}) \subset (I : (x_1, \dots, x_j)^{js'}) \subset (I : (x_1, \dots, x_j)^\infty),\end{aligned}$$

and the reverse inclusion is obvious.

Increasing s if necessary, we may assume that x_j does not divide m. It follows from the condition of Theorem 15.23 that $x_i^{s'} m \in I$ as required.

Suppose now that char $k = 0$. If $x_j^s m \in I$ then by the characterization of Theorem 15.23, any monomial $n = x_1^{s_1}, \cdots, x_j^{s_j}$ with $\sum_u s_u = s$ satisfies $nm \in I$, so $(I : x_j^s) \subset (I : (x_1, \dots, x_j)^s)$. Again, the reverse inclusion is obvious.

Corollary 15.25. *If I is a Borel-fixed ideal in S, and P is an associated prime of I, then $P = (x_1, \dots, x_j)$ for some j. If $Q = (x_1, \dots, x_t)$ is a maximal associated prime, then $x_{t+1}, \dots, x_r$ (in any order) is a maximal (S/I)-regular sequence in $(x_1, \dots, x_r)$.*

Proof. Since I is a monomial ideal, every associated prime of I is generated by a set of variables. Suppose that j is the largest index such that $x_j \in P$; we must show that $x_i \in P$ for $i < j$. Since P is an associated prime we may write $P = (I : f)$ for some polynomial f. Since $x_j f \in I$, it follows from Proposition 15.24 that $x_i^s f \in I$ for some s. Thus $x_i^s \in P$, so $x_i \in P$ as required.

If $Q = (x_1, \dots, x_t)$ is a maximal associated prime, then the variables $x_{t+1}, \dots, x_r$ cannot appear in the minimal generators of I. Thus $x_{t+1}, \dots, x_r$ (in any order) is a regular sequence mod I. Since Q is associated to I, there is a monomial $m \notin I$ such that $Qm \subset I$. Since the generators of I do not involve $x_{t+1}, \dots, x_r$, we may factor these variables out of m, and assume that $m \notin I + (x_{t+1}, \dots, x_r)$. It follows that $Q + (x_{t+1}, \dots, x_r) = (x_1, \dots, x_r)$ is associated to $I + (x_{t+1}, \dots, x_r)$, so $x_{t+1}, \dots, x_r$ is a maximal (S/I)-regular sequence in $(x_1, \dots, x_r)$.

For an analysis of these ideas and a different proof of Corollary 15.25, see Exercises 15.22 and 15.23.

15.10 Applications

We now apply these methods to the problems mentioned at the beginning of this section.

15.10.1 Ideal Membership

Given generators for an ideal $I \subset S$, determine a vector space basis for S/I, and given a polynomial f, compute its image in S/I in terms of this basis. If $f \in I$ (that is, if the image is 0) compute an expression for f as a linear combination of the generators of I.

This problem is solved by Theorem 15.3 and the division algorithm: Choose a monomial order on S, and from the original generators $f_1, \dots, f_s$ of I, compute a Gröbner basis $g_1, \dots, g_t$ for I. The set of monomials not in $\text{in}(I)$, that is, not divisible by any one of the $\text{in}(g_i)$, is a basis for S/I. The remainder of any $f \in S$ on division by $g_1, \dots, g_t$ has no monomials in $\text{in}(I)$ and is thus the unique expression for the image of f in terms of this basis.

If $f \in I$, then the division process exhibits f as a linear combination of the generators g_i. Since the algorithm that produces the g_i exhibits them as linear combinations of the original f_j, we are done.

For a generalization and a more formal treatment of the second part, see the application to syzygies.

15.10.2 Hilbert Function and Polynomial

Following Hilbert, we could deduce the Hilbert function or polynomial of a graded module from a graded free resolution for the module, computed with the algorithms above. However, this is extremely inefficient, and better schemes are based on the following fundamental result of Macaulay (1927). This theorem was the reason for Macaulay's introduction of monomial orders, and is thus historically at the very root of the material in this chapter.

Theorem 15.26. *Let P be a finitely generated, graded S-module, given by generators and relations as $P = F/M$, where F is a free module with a homogeneous basis and M is a submodule generated by homogeneous elements. The Hilbert function of P is the same as the Hilbert function of $F/\operatorname{in}(M)$.*

Proof. Let B be the set of monomials not in $\operatorname{in}(M)$. Write F_d for the set of elements of degree d of F, and similarly for M, P, and B. Because P is graded, we have $P = \oplus_d P_d$, where $P_d = F_d/M_d$.

By Theorem 15.3, B maps to a (vector space) basis for P, so B_d maps to a basis for P_d. Thus $\dim_k P_d$ is the number of elements of B_d. Since the argument applies as well to $P' = F/\operatorname{in}(M)$, we are done. □

Theorem 15.26 shows that to compute the Hilbert function of an arbitrary module, it is enough to compute the Hilbert function of the quotient of a free module by a monomial submodule, and this we have already done in the section on monomials.

Macaulay's original application of Theorem 15.26 was to give a characterization of all possible Hilbert functions of ideals: By virtue of the theorem, it is enough to characterize the Hilbert functions of monomial ideals, and this leads to a complex but manageable combinatorial problem. See Stanley [1978] and Green [1989b].

15.10.3 Associated Graded Ring

Let $R = S/I$, and set $\mathfrak{m} = (x_1, \ldots, x_r)$. The associated graded ring $\operatorname{gr}_{\mathfrak{m}} R$ of R with respect to $\mathfrak{m}$ is significant geometrically, algebraically, and computationally. The geometric and algebraic significances have been explained in Chapter 4; the main computational significance comes from the fact that its Hilbert function is the same as that of R and is easier to compute, for instance by the method above. To understand $\operatorname{gr}_{\mathfrak{m}} R$ we must find a presentation of the form $\operatorname{gr}_{\mathfrak{m}} R = S/I'$, where I' is the homogeneous ideal consisting of the sum f_{bottom} of the monomials of lowest degree from each polynomial f in I. Our goal is thus to produce finitely many elements g_i of I such that I' is generated by the forms $g_{i\,\text{bottom}}$. Interestingly, in order to do this we need only compute the Gröbner basis of a homogeneous ideals! A similar idea will suffice to compute the associated graded module of any S-module with respect to $\mathfrak{m}$; see Exercise 15.36. Of course, we could also ask for the associated graded ring of R (or any S-module) with respect to an arbitrary ideal I. This can be done by using elimination theory; see Exercise 15.38.

Choose any set of generators $f_1, \ldots, f_s$ of I, and for each i let F_i be the **homogenization** of f_i with respect to a new variable x_0, that is,

$$F_i(x_0, x_1, \ldots, x_r) = x_0^{\deg f_i} f_i(x_1/x_0, \ldots, x_r/x_0).$$

Proposition 15.28. *With notation as above, let* $(G_1, \ldots, G_t)$ *be a Gröbner basis of the ideal* $(F_1, \ldots, F_s)$ *with respect to any monomial order on* $S[x_0]$ *that refines the partial order by degree in* x_0*. If we set* $G_i(1, x_1, \ldots, x_r) = g_i(x_1, \ldots, x_r) \in S$*, then* $I' = (g_{1\ \text{bottom}}, \ldots, g_{t\ \text{bottom}})$.

Proof. Suppose $g \in I$; we must show that g_{bottom} is a linear combination of the $g_{i\,\text{bottom}}$. Write $g = \sum p_i f_i$. If G, P_i, and F_i are the homogenizations of g, p_i, and f_i, respectively, then for some integers a, b we have

$$x_0^a G = \sum P_i F_i \in (F_1, \ldots, F_s)$$
$$x_0^a G = x_0^b g_{\text{bottom}} + (\text{terms of lower degree in } x_0).$$

Because the G_i form a Gröbner basis for $(F_1, \ldots, F_s)$, there is a standard expression

$$x_0^a G = \sum Q_i G_i, \ \text{in}(Q_i G_i) \leq \text{in}(x_0^a G).$$

In particular, the degree in x_0 of $Q_i G_i \leq b$ for each i.

It follows that $x_0^b g_{\text{bottom}}$ is the sum of the products of the terms of highest degree in x_0 in some of the Q_i and G_i. Setting $x_0 = 1$, we see that g_{bottom} itself is the sum of the products of the terms of lowest degree in those $Q_i(1, x_1, \ldots, x_r)$ and $G_i(1, x_1, \ldots, x_r)$; that is, it is a linear combination of the $g_{i\ \text{bottom}}$, as claimed.

15.10.4 Elimination

Given an ideal $I \subset S[y_1, \ldots, y_s]$, we wish to compute $J = I \cap S$. The name "elimination" comes from thinking of the generators of I as a system of equations in x_i and y_j from which one wants to eliminate the variables y_j. We have already discussed a part of elimination theory in Chapter 14.

To eliminate variables using Gröbner bases, one uses an order on $T = k[x_1, \ldots, x_r, y_1, \ldots, y_s]$ satisfying:

$$\text{If } f \in T \text{ and } \text{in}(f) \in S, \text{ then } f \in S.$$

An order with this property is called an **elimination order (with respect to $y_1, \ldots, y_s$).**

Examples.

1. The simplest way to make an elimination order is to take the partial order by total degree in $y_1, \ldots, y_s$, refined by any monomial order; in practice it is often most efficient to take reverse lexicographic order as the second order.

2. Lexicographic order is an elimination order with respect to every initial subset of the variables.

To find $J = S \cap I$ we need only compute a Gröbner basis with respect to an elimination order:

Proposition 15.29. *Let $>$ be a monomial order on $T = S[y_1, \ldots, y_s] = k[x_1, \ldots, x_r, y_1, \ldots, y_s]$, and suppose that $>$ is an elimination order with respect to the variables $y_1, \ldots, y_s$. If $I \subset T$ is an ideal, then with respect to the monomial order on S obtained by restricting $>$, we have*

$$\mathrm{in}(I \cap S) = \mathrm{in}(I) \cap S.$$

Further, if $g_1, \ldots, g_t$ is a Gröbner basis for I, and $g_1, \ldots, g_u$ are those g_i that do not involve the variables y_i, then $g_1, \ldots, g_u$ form a Gröbner basis in S for $I \cap S$.

Proof. Let $J = I \cap S$. Clearly, $\mathrm{in}(J) \subset \mathrm{in}(I) \cap S$. We will show that the $\mathrm{in}(g_i)$ for $i \leq u$ generate $\mathrm{in}(I) \cap S$. By Lemma 15.5, this will prove both statements.

Suppose $m \in \mathrm{in}(I) \cap S$ is a monomial. Since $g_1, \ldots, g_t$ form a Gröbner basis, m is a multiple of $\mathrm{in}(g_i)$ for some $i \leq t$. Because $m \in S$, we must have $\mathrm{in}(g_i) \in S$, so $g_i \in S$ whence $i \leq u$ as required.

There is an analogue of Proposition 15.29 for submodules of a free module: If $M \subset F := \oplus Te_i$, then it gives us a way to construct $M \cap \oplus Se_i$. See Exercise 15.37.

One of the most frequent applications of elimination is solving the following problem:

Find the equations satisfied by given elements of an affine ring. (Geometrically: Find the closure of the image of an algebraic set under a morphism.)

Let $R = k[y_1, \ldots, y_s]/K$ for some ideal K, and let $f_1, \ldots, f_r$ be elements of R. Define a map

$$\varphi : S = k[x_1, \ldots, x_r] \to R; \qquad x_i \mapsto f_i.$$

We wish to find $\ker \varphi$. Geometrically, this is the ideal defining the Zariski closure of the image of the algebraic set corresponding to K under the map corresponding to $f_1, \ldots, f_r$.

To this end, set $Q = k[y_1, \ldots, y_s]$, and consider the ring $T = k[x_1, \ldots, x_r, y_1, \ldots, y_s]$. For each i, let $F_i \in Q$ be a polynomial that maps to $f_i \in R$. Regarding the F_i as elements of T, let $I \subset T$ be the ideal

$$I = KT + (F_1 - x_1, \ldots, F_r - x_r).$$

Proposition 15.30. $\ker \varphi = I \cap S$.

Proof. Consider the map $\bar{\varphi} : T \to Q$ sending $x_i \mapsto F_i$. The ideal $J := (F_1 - x_1, \ldots, F_r - x_r)$ is obviously contained in $\ker \bar{\varphi}$. We claim that $J = \ker \bar{\varphi}$.

Indeed, it is clear that $J \subset \ker \bar{\varphi}$, and the reverse inclusion follows because each x_i is equal to a polynomial in the $y_j \bmod J$.

It follows that the kernel of the composite map $T \to Q \to R$ is I, so the kernel of the composite $S \hookrightarrow T \to Q \to R$ is $S \cap I$, as claimed. □

If K and the f_i are homogeneous, then $\ker \varphi$ will be a homogeneous ideal too if we take the variables x_i to have the same degrees as the f_i. If, for example, all the f_i were of the same degree, then we could change gradings to give the x_i all degree 1, and the ideal $\ker \varphi$ would remain homogeneous. Here we are computing the equations for the projective variety that is the image of $V(K)$ under the map corresponding to φ—in this case the image is already closed, by the main theorem of elimination theory, Theorem 14.1.

15.10.5 Projective Closure and Ideal at Infinity

Given an algebraic set $V \subset \mathbf{A}^r$, we wish to compute the ideal I' of the closure $\bar{V}$ of V in $\mathbf{P}^s \times \mathbf{A}^{r-s}$ and the ideal I_∞ of the intersection of $\bar{V}$ with the **hyperplane at infinity** $\mathbf{P}^{s-1} \times \mathbf{A}^{r-s} \subset \mathbf{P}^s \times \mathbf{A}^{r-s}$.

To describe them it is convenient to introduce the term ***s*-degree** to denote the degree of a polynomial with respect to the first s variables of the polynomial ring S. The ***s*-homogenization** of a polynomial $g(x_1, \ldots, x_r)$ of s-degree d, with respect to a new variable x_0, is then defined as

$$g'(x_0, x_1, \ldots, x_r) := x_0^d g(x_1/x_0, \ldots, x_s/x_0, x_{s+1}, \ldots, x_r).$$

Less formally, p' is the sum of the terms of p, each multiplied by a power of x_0 to bring it up to s-degree d.

If $f \in S[x_0] = k[x_0, x_1, \ldots, x_r]$ is a form homogeneous with respect to $x_0, \ldots, x_s$, and $W \subset \mathbf{P}^s \times \mathbf{A}^{r-s}$ is the corresponding hypersurface, then $W \cap (\mathbf{A}^s \times \mathbf{A}^{r-s}) = W \cap \mathbf{A}^r$ is defined by the equation $f(1, x_1, \ldots, x_r) = 0$. It follows easily that I' is the set of elements in the preimage of I under the map

$$\varphi : S[x_0] \to S; \qquad x_0 \mapsto 1$$

that are homogeneous in the variables $x_0, \ldots, x_s$. Equivalently, I' is the ideal generated by the s-homogenizations g' of all the elements $g \in I$. From this second description and the fact that the hyperplane at infinity has equation $x_0 = 0$, we see that we may write $I_\infty = \{g_\infty | g \in I\}$, where g_∞ denotes the sum of all those terms of g with maximal s-degree.

The problem is that these descriptions of I' and I_∞ involve infinitely many polynomials. (It is easy to show that if $\{h_i\} \subset I$ is any set of elements such that the set $(\{h_i)_\infty\}$ generates I_∞, then I' is generated by the set of s-homogenizations of the h_i, but this still does not solve the problem.) The following result shows how to compute both I' and I_∞ in finite terms, using Gröbner bases.

Proposition 15.31. *With notation as above, suppose that $>$ is a monomial order on S refining the order by degree in $x_1, \ldots, x_s$. If $g_1, \ldots, g_t$ is a Gröbner basis of I with respect to $>$, then*

a. $\text{in}(I_\infty) = \text{in}(I)$ *and* $\{(g_1)_\infty, \ldots, (g_t)_\infty\}$ *is a Gröbner basis for* I_∞.

b. $\text{in}(I') = \text{in}(I)$ *and* $\{g'_1, \ldots, g'_t\}$ *is a Gröbner basis of* I'.

Proof.

a. We have $\text{in}(g_\infty) = \text{in}(g)$ for any $g \in S$ by our choice of order. It follows that $\text{in}(I_\infty) = \text{in}(I')$, and this ideal is generated by the $\text{in}(g'_i) = \text{in}(g_i)$.

b. Again, we have $\text{in}(g') = \text{in}(g)$ for any $g \in S$ by our choice of order. As in part a, it follows that $\text{in}(I') = \text{in}(I)$, and this ideal is generated by the $\text{in}(g'_i) = \text{in}(g_i)$. □

15.10.6 Saturation

If M is a submodule of a free S-module F and J is an ideal of S, we define

$$(M : J) = \{f \in F | fJ \subset M\} \subset F$$
$$(M : J^\infty) = \bigcup_{d=1}^{\infty} (M : J^d) \subset F.$$

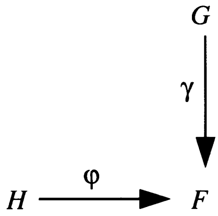
The submodule $(M : J^\infty)$ is called the **saturation of M with respect to J**. Saturations arise in the theory of primary decomposition and in local cohomology theory. They can also be used for finding projective closures—see Exercise 15.40.

In the section on applications of syzygies, we will see that we can compute $(M : J^d)$. We could compute these one at a time, increasing d until we obtained $(M : J^d) = (M : J^{d+1})$ (which must happen eventually because S is Noetherian). For this value of d we have $(M : J^\infty) = (M : J^d)$, and this is a rather practical method in many cases. However, part a of Exercise 15.41, with $d = \infty$, shows that if J is the ideal generated by a single variable, then the problem can be solved using a single Gröbner basis computation with respect to a suitable order. The general case can easily be reduced to the special case, using the other ideas in Exercise 15.41.

15.10.7 Lifting Homomorphisms

The following generalization of the ideal membership problem is central to many constructions involving maps of modules and homological algebra. The application below to the kernels of maps of modules is an example.

Let F, G, and H be free S-modules with base, and suppose we are given maps such that $\text{im}\,\gamma \subset \text{im}\,\varphi$. We would like to construct a "lift" $\psi : G \to H$ such that $\varphi\psi = \gamma$.



To this end, we choose a monomial order on F. Write $h_1, \ldots, h_s$ for the images under φ of the basis vectors of H. Using Buchberger's algorithm, we may find a Gröbner basis $h'_1, \ldots, h'_t$ for $\operatorname{im}\varphi$. Let

$$\varphi' : H' = \oplus S e_i \to F; \qquad e_i \mapsto h'_i$$

be the corresponding map. Buchberger's algorithm produces at the same time an expression for each h'_i in terms of the h_j; that is, a "change of basis map" $\alpha : H' \to H$ such that $\varphi' = \varphi\alpha$. For each basis vector $\varepsilon_i \in G$, we use the division algorithm to find an expression $\gamma(\varepsilon_i) = \sum p_i h'_i$. We may define a map

$$\psi' : G \to H'; \qquad \varepsilon_i \mapsto \sum p_i e_i,$$

so that $\varphi'\psi' = \gamma$ (see Figure 15.7). It follows that $\psi = \alpha\psi'$ is the desired lifting.

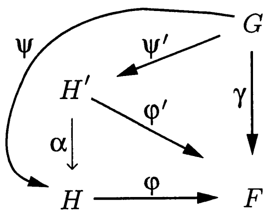


FIGURE 15.7.

15.10.8 *Syzygies and Constructive Module Theory*

A module may be determined in many ways by giving its properties. By contrast, we will say that we have **constructed** a module P only if we can give generators and relations for it—that is, if we can write it as F/M where F is a free module and M is a submodule generated by explicitly given elements of F, or equivalently as the cokernel of a map $\varphi : G \to F$ of free modules with image M. Since we can compute $\ker\varphi$, we may also regard the submodule M as having been constructed. If we have constructed a module $P = F/M$, and have specified a submodule $P' \subset P$ by giving a set of generators for it as the images of given elements of F, then it is clear that we can construct the quotient P/P'; we simply adjoin the new elements of F to the list of relations for P. It is not quite so obvious that

we can find the relations for the submodule P', but this will follow from the pullback construction we are about to describe (Exercise 15.45). In the remainder of this section we will make a few of the central operations on modules constructive in this sense. The list here could be prolonged very greatly.

8.a Pullbacks, Intersections, and Annihilators: If $\varphi : G \to F$ and $\psi : H \to F$ are maps of free modules, it is often useful to find the "pullback" of φ and ψ, by which we mean the submodule of $G \oplus H$ consisting of elements (g, h) such that $\varphi(g) = \psi(h)$. The pullback is the kernel of $(\varphi, -\psi) : G \oplus H \to F$. Since we are able to compute syzygies (Theorem 15.10), we can find a free module PB mapping onto the kernel—that is, we can find a set of generators for the pullback. We will often need the projection to one of the factors; we will write $\pi_G : PB \to G$ and $\pi_H : PB \to H$ for the compositions of $PB \to G \oplus H$ with the projections. We get the following commutative diagram.

$$\begin{array}{ccc} PB & \xrightarrow{\pi_G} & G \\ {\scriptstyle \pi_H} \downarrow & & \downarrow {\scriptstyle \varphi} \\ H & \xrightarrow[\psi]{} & F \end{array}$$

This gives us a way to compute the intersection of two submodules of F, the images of φ and ψ, say. The intersection is simply the image of $\varphi\pi_G = \psi\pi_H$. See Exercises 15.42 and 15.43 for other constructions of intersections, and Exercises 15.45 and 15.46 for further uses of pullbacks.

8.b The Kernel of a Map between Arbitrary Modules: Given S-modules P and Q by means of free presentations

$$G_P \overset{\kappa_P}{\to} F_P \to P \to 0$$

$$G_Q \overset{\kappa_Q}{\to} F_Q \to Q \to 0,$$

and given a homomorphism $P \to Q$ presented as a map $\varphi : F_P \to F_Q$ taking the image of G_P into the image of G_Q, we may construct the kernel of the map $P \to Q$ induced by φ:

Proposition 15.32. *With notation as above, let F_K be a free module mapping onto the pullback of κ_Q, φ, and let $\psi : F_K \to F_P$ be the composite*

$$F_K \to G_Q \oplus F_P \to F_P.$$

Let $\kappa'_K : G \to F_K$ be a map onto the kernel of ψ. Let $\varphi_1 : G_P \to G_Q$ be a map lifting the map $\varphi\kappa_P$ along κ_Q and let $\kappa''_K : G_P \to F_K$ be a lifting of the map $(-\varphi_1, \kappa_P) : G_P \to G_Q \oplus F_P$, which maps into the kernel of (κ_Q, φ). Then

$$\kappa_K = (\kappa'_K, \kappa''_K) : G_P \oplus G \to F_K \to K \to 0$$

is a free presentation of the kernel of $P \to Q$, and the injection $K \subset P$ is defined by the map ψ.

We leave the proof to the reader (Exercise 15.44).

Hom, Ext, Tor, and all that: Much can be computed by putting together what we have already done. We give only some hints, and leave the working out of these constructions to the reader with sufficient background.

If P and Q are S-modules given by free presentations as above, then

$$\mathrm{Hom}_S(P, Q) = \ker(\mathrm{Hom}(F_P, Q) \to \mathrm{Hom}(G_P, Q)),$$

while $\mathrm{Hom}(F_P, Q)$ is a module with free presentation

$$\mathrm{Hom}(F_P, G_Q) \to \mathrm{Hom}(F_P, F_Q) \to \mathrm{Hom}(F_P, Q) \to 0,$$

and similarly for $\mathrm{Hom}(G_P, Q)$.

Once we can compute free resolutions, Hom, and kernels, Ext is easy; and the same is true for Tor if we can compute tensor products. But tensor products are elementary (that is, one doesn't need to solve equations) because, for example, the tensor product of P and Q is presented as

$$F_P \otimes G_Q \oplus F_Q \otimes G_P \to F_P \otimes F_Q \to P \otimes Q \to 0,$$

by the right-exactness of tensor products.

Multiplicities (in the sense of Samuel or Serre) can be computed from computations of Tor; those in the sense of Vogel can also be found, using the computation of saturations.

The cohomology of coherent sheaves can be handled from Ext either using duality theory or directly, since the usual expression for the cohomology as the limit of certain Ext groups actually converges, in each degree, in a predictable, finite number of steps. More generally, local cohomology can be approximated. The interested reader may find details of these and other constructions in Vasconcelos [in preparation].

15.10.9 What's Left?

Many further things can be done with Gröbner bases well enough to have been implemented on computers (for example, in the program *Macaulay*). In this category fall, for instance,

- Finding syzygies over factor rings
- Computing the radical of an ideal
- Primary decomposition (the first algorithm was by Grete Hermann, a student of Emmy Noether [1926]; for recent work, see Seidenberg [1984], Gianni, Trager, and Zacharias [1988], and Eisenbud, Huneke, and Vasconcelos [1992])

Other algorithms are known but not implemented for various reasons. A few examples from many:

- Normalization of a ring (this was studied by Seidenberg [1974]; for recent work see Vasconcelos [1992])
- Flattening stratifications

There are also plenty of problems where no algorithms are known as of this writing. Again a few examples:

- Decide whether a module can be written as a direct sum of submodules nontrivially; if so, decompose it. For example, decide whether a projective module is free.
- Decide whether two varieties are in the same component of the Hilbert scheme.
- Compute the versal deformation of a factor ring S/I in the case that this is finite-dimensional.
- Decide the growth rate of the (infinite) free resolution of a module over a factor ring of S.
- Given generators for an ideal, decide whether a smaller number of generators can generate an ideal with the same radical; in particular, decide whether an algebraic variety is a "set theoretic complete intersection"—that is, set theoretically the intersection of c hypersurfaces, where c is the codimension. The leading open case is perhaps the ideal of the rational quartic curve in $\mathbf{P}^3$:

  $$(x_2^3 - x_1x_3^2, x_1x_2 - x_0x_3, x_1^3 - x_0^2x_2, x_0x_2^2 - x_1^2x_3) \subset k[x_0, \dots, x_3],$$

  which is the kernel of the map

  $$\begin{aligned} k[x_0, \dots, x_3] &\to k[s,t] \\ x_0 &\mapsto s^4 \\ x_1 &\mapsto s^3t \\ x_2 &\mapsto st^3 \\ x_3 &\mapsto t^4. \end{aligned}$$

  It is known that if the characteristic of k is positive, then this ideal has the same radical as an ideal generated by two elements (the elements depend on the characteristic). It is not known whether this is true in characteristic 0. See Jaffe [1989] for recent results and an exposition.

15.11 Exercises

In Exercises 15.1–15.6, m, n, m_i, and n_i denote monomials.

Exercise 15.1:* Solve the "elimination problem" in the monomial case: If $I = (m_1, \ldots, m_t) \in S$ and $s < r$, find $I \cap k[x_1, \ldots, x_s]$.

The next two (easy) exercises are used in the computation of Hilbert functions and polynomials.

Exercise 15.2:* Show that any monomial submodule of a free module $\oplus Se_i$ is a direct sum of modules of the form $I_i e_i$ with I_i a monomial ideal of S.

Exercise 15.3:* Let $I = (m_1, \ldots, m_t)$ be a monomial ideal, and let n be a monomial of S. Prove that the ideal

$$(I : n) = \{f \in S | fn \in I\}$$

is generated by the monomials $m_i/\text{GCD}(m_i, n)$.

Exercise 15.4: Show that if I is a monomial ideal, then the Hilbert function or polynomial of S/I can be computed as a sum of binomial coefficients by using the following "divide-and-conquer" strategy:

a. First, if I is generated by some number s of the r variables of S, then

$$H_{S/I}(\nu) = H_{k[x_1,\ldots,x_{r-s}]}(\nu) = \binom{r-s-1+\nu}{r-s-1}.$$

We can think of the binomial coefficient "combinatorially," so that it is 0 for all sufficiently small ν, in which case it is the Hilbert function, or as a polynomial in ν of degree$(r-s-1)$, in which case it is the Hilbert polynomial.

b. If I is not generated by such a subset of variables, let $n \in S$ be any monomial properly dividing one of the minimal generators of I, and let d be the degree of n. Write $J := (I : n)$. Show that there is an exact sequence of graded modules and degree-0 maps

$$0 \to S/J(-d) \xrightarrow{\varphi} S/I \to S/(I, n) \to 0$$

and thus

$$H_{S/I}(\nu) = H_{S/J}(\nu - d) + H_{S/(I,n)}(\nu).$$

It is an open problem to determine the most efficient choice for n, but an obvious idea is to take it to be "half" the "largest" monomial among the generators of I.

Exercise 15.5: Let $I = (x_1x_3, x_1x_4, x_2x_4)$. Compute the Hilbert function and the Hilbert polynomial of I.

Exercise 15.6: For each of the following ideals, compute a minimal set of divided Koszul relations that generates the syzygies:

a. $(x_1^{34}x_2^7, x_1^{23}x_2^{19})$

b. (x_1, x_2, x_3)

c. (x_1x_2, x_1x_3, x_2x_3)

Exercise 15.7:* Let $I_1 = (m_1, \ldots, m_s)$ and $I_2 = (n_1, \ldots, n_t)$ be monomial ideals of S. Show that $I_1 \cap I_2$ is generated by the elements LCM (m_i, n_j). When is this equal to the ideal I_1I_2?

Exercise 15.8:

a. If F is a free module with basis, $M \subset F$ is any monomial submodule, and $>$ is any monomial order on F, then $\mathrm{in}_>(M) = M$.

b. For any submodule M show that $\mathrm{in}_>(M)$ is spanned as a vector space by the elements $\{\mathrm{in}_>(f) | f \in M\}$; that is, we do not need to impose the condition that $\mathrm{in}_>(M)$ is a submodule.

Exercise 15.9: If $I \subset S$ is a homogeneous ideal, show that $\mathrm{in}_>(I)$ is generated by the monomials $\{\mathrm{in}_>(f) | f \in I \text{ is a homogeneous polynomial}\}$.

Exercise 15.10: Show that the following properties characterize the orders $>_{\mathrm{lex}}$, $>_{\mathrm{hlex}}$, and $>_{\mathrm{rlex}}$ among monomial orders on S:

a. If $\mathrm{in}_{\mathrm{lex}}(f) \in k[x_s, \ldots, x_r]$ for some s, then $f \in k[x_s, \ldots, x_r]$.

b. $>_{\mathrm{hlex}}$ refines the order by total degree; and if f is homogeneous with $\mathrm{in}_{\mathrm{hlex}}(f) \in k[x_s, \ldots, x_r]$ for some s, then $f \in k[x_s, \ldots, x_r]$.

c. $>_{\mathrm{rlex}}$ refines the order by total degree; and if f is homogeneous with $\mathrm{in}_{\mathrm{rlex}}(f) \in (x_s, \ldots, x_r)$ for some s, then $f \in (x_s, \ldots, x_r)$.

More generally, suppose F is a free module with basis over S having a reverse lexicographic monomial order, and $f \in F$. If $\mathrm{in}_{\mathrm{rlex}}(f) \in (x_s, \ldots, x_r)F$ for some s, then $f \in (x_s, \ldots, x_r)F$.

Exercise 15.11: Given a monomial order $<$ on S, define the **positive cone** $P_< \subset \mathbf{Z}^r$ of $>$ to be the set of differences $a - b$ such that a, b are vectors of nonnegative integers and (in multiindex notation for monomials) $x^a > x^b$. Show that P is a convex cone in the sense that

$$u, v \in P_< \Rightarrow pu + qv \in P_< \text{ whenever } 0 \le p, q \in \mathbf{Q} \text{ and } pu + qv \in \mathbf{Z}^r,$$

and is even strictly convex in the sense that

$$u \in P_< \Rightarrow -u \notin P_<.$$

Exercise 15.12:* Let $>$ be a monomial order on S, and suppose that m_i, n_i are monomials such that $m_i > n_i$ for $i = 1, \ldots, t$. Show that there is an integral weight order defined by some $\lambda : \mathbf{Z}^r \to \mathbf{Z}$ such that λ is compatible with $>$ and $m_i >_\lambda n_i$ for $i = 1, \ldots, t$ (Bayer [1982]).

Exercise 15.13:* Show that every monomial order on S is a lexicographic product of at most r weight orders λ_i (Robbiano, [1986]).

Exercise 15.14: Let F be a free module with basis, and fix a monomial order on F. Suppose that $g_1, \ldots, g_t \in M \subset F$.

a. Prove that if $\text{in}(M)$ is generated by $\text{in}(g_1), \ldots, \text{in}(g_s)$, then $g_1, \ldots, g_s$ is also a Gröbner basis for M. If $\text{in}(g_1), \ldots, \text{in}(g_s)$ is a minimal set of generators for $\text{in}(M)$, then $g_1, ldots, g_s$ is called a **minimal Gröbner basis** of M.

b. Show that there exists a Gröbner basis $h_1, \ldots, h_s$ for M with the properties

   i. $\text{in}(h_i)$ is a monomial (that is, the coefficient from k is 1),

   ii. $\text{in}(h_i)$ does not divide any term of h_j for $i \neq j$.

Show that for such a Gröbner basis, the elements $\text{in}(h_i)$ are the minimal generators of $\text{in}(M)$. Show that if $g_1, \ldots, g_s$ also has properties i and ii, then $\{g_i\} = \{h_i\}$. The Gröbner basis $h_1, \ldots, h_s$ is called **the reduced Gröbner basis** of M.

Exercise 15.15 (Gordan's Proof of the Hilbert Basis Theorem): Gordan, initially shocked by Hilbert's proof of the finite generation of certain rings of invariants by means of the basis theorem, recovered quickly and gave his own, simplified proof in [1900]. This proof represents an early (the earliest?) use of the idea of an "initial" ideal of monomials associated to an ideal in a polynomial ring. Here is a proof of the Hilbert basis theorem, in the spirit of Gordan. (Gordan needed only a special case, and thus proved only a special case, though his argument works generally. It can even be extended to give a proof of the form of the basis theorem saying that if R is a Noetherian ring then $R[x]$ is too.)

a. Give a combinatorial proof that any set of monomials of $S = k[x_1, \ldots, x_r]$ has only finitely many minimal elements in the partial order by divisibility. (This part is sometimes called "Dickson's lemma".) In particular, every monomial ideal is finitely generated.

b. By part a, any ideal in S has a finite Gröbner basis (with respect to any given monomial order). Deduce that S is Noetherian.

Exercise 15.16:* Show that the division algorithm still terminates if at each stage we simply choose some monomial of f_t' divisible by some $\mathrm{in}(g_i)$, instead of the greatest such. This gives a still more indeterminate version of the division algorithm, which works just as well for the purposes of this chapter as the one we gave earlier.

Exercise 15.17 (Characterization of determinate division): Suppose that $f = \sum m_u g_{s_u} + f'$ is the standard expression for f with respect to $g_1, \ldots, g_t$ produced by the determinate division algorithm. If we take h_v to be the sum of all the terms m_u such that $s_u = v$, we may rewrite this expression as

$$f = \sum h_v g_v + f'.$$

Show that this is the unique such expression for which the monomials of h_v lie in the set of monomials n of S such that

$$n\,\mathrm{in}(g_v) \notin (\mathrm{in}(g_1), \ldots, \mathrm{in}(g_{v-1}))$$

and the monomials of f' do not lie in $(\mathrm{in}(g_1), \ldots, \mathrm{in}(g_t))$.

Exercise 15.18:* Prove that with notation as in Theorem 15.10, $\ker \varphi$ is generated by any set of τ_{ij} such that the corresponding σ_{ij} generate the syzygies on the elements $\mathrm{in}(g_i)$.

The following two results of Buchberger sometimes help to speed up the process of computing a Gröbner basis.

Exercise 15.19: Imitate the proof of Theorem 15.8 to show that in applying Buchberger's criterion it is enough to check any subset of pairs i, j such that the corresponding σ_{ij} generate all the syzygies on the elements $\mathrm{in}(g_i)$.

Exercise 15.20:* With notation as in Algorithm 15.9, suppose $F = S$. Show that if $\mathrm{in}(g_i)$ and $\mathrm{in}(g_j)$ are relatively prime, then the division algorithm can be carried out so that the remainder on division of $m_{ji}g_i - m_{ij}g_j$ by g_i and g_j is 0, and thus the remainder on division of $m_{ji}g_i - m_{ij}g_j$ by $(g_1, \ldots, g_t)$ may be taken to be 0. Thus such syzygies of the $\mathrm{in}(g_i)$ may be ignored in computing a Gröbner basis. (This is a case where it is good to have an indeterminate division algorithm!)

Exercise 15.21:* Some plausible-sounding variations on Proposition 15.15 are *false*. For simplicity we take the case $F = S$. Let $I \subset S$ be an ideal, and choose a monomial order on S. Find an example of a sequence of elements $h_1, \ldots, h_u \in S$ such that $h_1, \ldots, h_u$ is a regular sequence on $S/\,\mathrm{in}(I)$,

and $\mathrm{in}(h_1), \ldots, \mathrm{in}(h_u)$ is a regular sequence on S/I, but $h_1, \ldots, h_u$ is not a regular sequence on S/I.

Exercise 15.22: If I is an ideal of a Noetherian ring S, and $x, y \in S$, show that the following are equivalent:

1. $(I : y^\infty) = (I : (x,y)^\infty)$.
2. Every associated prime of I that contains y also contains x.

Exercise 15.23: Prove that any closure of an orbit of $\mathcal{B}$ on k^r is, for some i, the subspace spanned by the last i basis elements. Use this to give another proof of Corollary 15.25.

Exercise 15.24 (Bruhat):* If g is an $r \times r$ matrix, then the **principal minor** of order $s \leq r$ is the determinant of the upper-left $s \times s$ submatrix of g; that is, if $g = (g_{ij})_{1 \leq i,j \leq r}$ then the principal minor of order s is $\det((g_{ij})_{1 \leq i,j \leq s})$. If $\mathcal{U}$ is the set of upper triangular $r \times r$ matrices with ones on the diagonal, and $\mathcal{B}'$ is the set of invertible lower triangular matrices, show that $\mathcal{B}'\mathcal{U}$ is the set of invertible matrices whose principal minors are all nonzero. In particular, $\mathcal{B}'\mathcal{U}$ is a Zariski open and—if k is infinite—dense subset of $\mathcal{G}$. (In fact, $\mathcal{B}'\mathcal{U}$ is the "big cell" in the Bruhat decomposition of $\mathcal{G}$; see Humphreys [1975] or Fulton and Harris [1991] for more of the story.)

Exercise 15.25: With notation as in Theorem 15.17 show that if $g_1, \ldots, g_t \in I$ are chosen so that $\mathrm{in}_>(g_1), \ldots, \mathrm{in}_>(g_t)$ generate $\mathrm{in}_>(I)$, or even so that $\mathrm{in}_\lambda(g_1), \ldots, \mathrm{in}_\lambda(g_t)$ generate $\mathrm{in}_\lambda(I)$, then $\tilde{g}_1, \ldots, \tilde{g}_t$ generate $\tilde{I}$.

Exercise 15.26: Let $>$ be a monomial order on S, and let T be the subring of the quotient field of S generated by all the fractions m/n, with m and n monomials of S such that $m \geq n$ (we consider $m > 1$ for any nontrivial monomial, so T contains the polynomial ring S).

a.* Show that $\{m/n | m, n \in S \text{ are monomials and } m > n\}$ generates a proper ideal J of T, and that the quotient T/J is k. Show that

$$S \subset T \subset S[x_1^{-1}, \ldots, x_r^{-1}] = T[x_1^{-1}, \ldots, x_r^{-1}].$$

Show that T need not be Noetherian.

We will consider $\tilde{S} := T \otimes_k S$ as a flat family of algebras over T (it is flat because S is flat—indeed, free—as a k-module). For convenience of notation, we think of T as coming from a polynomial ring in a different set of variables, $y_1, \ldots, y_r$. With this notation, the fractions

$$x^\alpha y^\beta / y^\gamma \quad \text{with } y^\beta \geq y^\gamma$$

form a k-basis for $\tilde{S}$.

For any polynomial $g(x_1, \ldots, x_r) \in S$, with initial monomial x^α, let $\tilde{g} \in \tilde{S}$ be defined as

$$\tilde{g} = y^\alpha g(x_1/y_1, \ldots, x_r/y_r),$$

which is in $\tilde{S}$ precisely because all the monomials of g are $\leq x^\alpha$. For any ideal I of S, let $\tilde{I} \subset \tilde{S}$ be the ideal generated by all $\tilde{g}$ with $g \in I$.

b. Show that $T[y_1^{-1}, \ldots, y_r^{-1}] \otimes_T \tilde{S}/\tilde{I} \cong T[y_1^{-1}, \ldots, y_r^{-1}] \otimes_T S/I$ while $T/J \otimes_T \tilde{S}/\tilde{I} \cong S/\operatorname{in}_>(I)$.

c. Show that $\tilde{S}/\tilde{I}$ is flat over T by showing that it is free on the monomials in $x_1, \ldots, x_r$ not in $\operatorname{in}_>(I)$.

Exercise 15.27 (The simplest nontrivial syzygy computation):* Take $g_1 = x^2$, $g_2 = y^2$, $g_3 = xy + yz \in k[x, y, z]$. Find a Gröbner basis and syzygies using the reverse lexicographic order, and $x > y > z$.

Exercise 15.28 (Five points in $\mathbf{P}^3$):* Find a minimal free resolution of the ideal

$$I = (x_0^2 - x_2x_3, x_0x_1 - x_3^2, x_0x_2 - x_1^2, x_1x_3 - x_2^2, x_0x_3 - x_1x_2)$$

in the polynomial ring $S = k[x_0, \ldots, x_3]$ (this is the ideal of five points in $\mathbf{P}^3$).

Exercise 15.29:* Let $M = (x^2, txy + y^3) \subset k[t, x, y]$. Compute a Gröbner basis with respect to reverse lexicographic order using $t > x > y$.

Exercise 15.30: Using the result of Exercise 15.29, find a presentation for the associated graded ring of $k[x, y]/(x^2, xy + y^3)$ with respect to the ideal (x, y).

Exercise 15.31: Let $I = (x_1x_3 - x_2^2, x_1x_4 - x_2x_3, x_2x_4 - x_3^2)$ be the ideal of 2×2 minors of the matrix

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix}$$

and let I' be the ideal of minors of the matrix

$$\begin{pmatrix} x_2 & x_1 & x_3 \\ x_1 & x_3 & x_4 \end{pmatrix}$$

obtained by interchanging x_1 and x_2. Find Gröbner bases for I and I' with respect to the reverse lexicographic order on the monomials.

Exercise 15.32: Let I be the ideal of Exercise 15.31. Is $x_2^4 \in I$?

Exercise 15.33: Let R be the ring $k[x_{11}, x_{12}, x_{21}, x_{22}]$ and let $x =$

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$$

be the generic 2×2 matrix over R. Let I be the ideal generated by the four entries of the matrix X^2, so that $I = (x_{11}^2 + x_{12}x_{21}, \ldots)$. If the polynomials of I vanish when we substitute the entries a_{ij} of some 2×2 matrix A over k, then evidently $A^2 = 0$—that is, A is nilpotent. It follows from the Nullstellensatz that if $g(x_{11}, \ldots, x_{22})$ is any polynomial vanishing on 2×2 nilpotent matrices, then some power of g lies in I. The trace and determinant, $x_{11} + x_{22}$ and $x_{11}x_{22} - x_{12}x_{21}$ are such polynomials. Compute a Gröbner basis of I and use it and the division algorithm to decide which powers of the trace and determinant lie in I.

It is known that if X is a generic $n \times n$ matrix, then the coefficients of the characteristic polynomial of X (in the 2×2 case, the trace and determinant) generate the prime ideal corresponding to the variety of nilpotent $n \times n$ matrices, and one can ask in general what is the smallest integer d such that T^d belongs to the ideal of entries of X^n, where T is the trace $x_{11} + \cdots + x_{nn}$. Considering diagonal matrices it is easy to show that $d \geq n^2 - n + 1$. In fact, B. Mourrain has shown me a proof that $d = n^2 - n + 1$ using a "Sagbi base" for the ring of invariants under the conjugation action of $\mathrm{GL}(n, k)$, the ring generated by the coefficients of the characteristic polynomial. See Robbiano and Sweedler [1990] for the definitions.

Exercise 15.34 (The general submodule membership problem): Let P be any S-module, given by "generators and relations," that is, as $P = F/M$, where F is a free module with basis and $M = (f_1, \ldots, f_t)$ is a submodule of F. Let $Q \subset P$ be a submodule, given as the image of a submodule N of F. Generalize the idea in application 1 (Ideal Membership) to decide, for any element $p \in P$, whether or not $p \in Q$.

Exercise 15.35: Compute the Hilbert function and polynomial of the determinantal ideal I from Exercise 15.31.

Exercise 15.36:* Let P be any finitely generated S-module, and write $P = F/M$ with F a free S-module. Let $\mathfrak{m} = (x_1, \ldots, x_r)$. Use a technique analogous to that of Proposition 15.28 to construct a homogeneous submodule $M' \subset F$ (with F regarded as a graded module having all its generators in degree 0) such that $\mathrm{gr}_{\mathfrak{m}} M = F/M'$.

Exercise 15.37: Let $T = S[y_1, \ldots, y_s]$ and let F be a free T-module with basis e_i. Let $>$ be an elimination order on F with respect to the variables y_i. If $g_1, \ldots, g_t$ is a Gröbner basis in F, and $g_1, \ldots, g_s$ are those of the g_i that do not involve the variables y_i, show that $g_1, \ldots, g_s$ is a Gröbner basis in $F' = \oplus S e_i$ for $J = F' \cap (g_1, \ldots, g_t)$ with respect to the monomial order on F' gotten by restricting the given one from F.

Exercise 15.38: Show how to use elimination, via Proposition 15.30, to find presentations of the blowup algebra and associated graded ring of a ring S/I with respect to a given ideal $\mathfrak{m}$.

Exercise 15.39 (Nonhomogeneous Gröbner bases from homogeneous ones): Some computer algebra systems handle Gröbner bases only in the homogeneous case. The following shows that this is enough to compute Gröbner bases of arbitrary ideals of S.

Given a monomial order $>$ on S, extend it to $S[x_0]$ as follows: If m and n are monomials of S, define $mx_0^d > nx_0^e$ if $m > n$ or $m = n$ and $d > e$. Suppose that I is a (not necessarily homogeneous) ideal of S, and I' is any ideal of $S[x_0]$ that goes to I under the "specialization" map $S[x_0] \to S$ sending $x_0 \mapsto 1$ and $x_i \mapsto x_i$ for $i > 0$. Show that $\text{in}(I')$ goes to $\text{in}(I)$ under the specialization, and that any Gröbner basis of I' goes to a Gröbner basis of I under the specialization.

Exercise 15.40 (Projective closure by saturation): Let $I \subset S$ be an ideal, and let I' be the ideal of s-homogeneous elements (in the sense of the section on projective closures) in the preimage of I under the map

$$S[x_0] \to S; \qquad x_0 \mapsto 1.$$

If I'' is the ideal obtained by s-homogenizing the elements of some set of generators for I, show that

$$I' = (I'' : x_0^\infty).$$

Exercise 15.41 ((M:J) and (M:J^∞) in general): Suppose that F is a finitely generated free S-module and $M \subset F$ is a submodule. Let $J \subset S$ be any ideal. We wish to compute $(M : J)$ and $(M : J^\infty)$.

a. (Solution of the problem in case J is generated by a variable.) Suppose that $J = (x_r)$. Proposition 15.12b allows one to compute $(M : J)$ in this case. Show that Proposition 15.12 remains true if x_r is replaced by x_r^d for any $d \leq \infty$; the case $d = \infty$ gives a computation of $(M : J^\infty)$. This idea comes from Bayer [1982].

b. (Reduction to the case where J is a principal ideal.) Let $S' = S[y]$, where y is a new indeterminate, and regard S as a subring of S'. Let $M' = S' \otimes_S M \subset S' \otimes_S F$. Suppose $J = (f_1, \dots, f_t)$, and let $f = f_1 + yf_2 + \ldots + y^{t-1}f_t$. Show that $(M' : f) = S'(M : J)$, and thus $(M : J) = (M' : f) \cap F$, and deduce similar formulas for $(M : J^\infty)$. (One could also do this by introducing t new variables y_i, and using $f = \sum y_i f_i$. This is often less efficient computationally.)

c. (Reduction to the case where J is generated by a variable.) Suppose that $J = (f)$ is a principal ideal. Let $M' = S' \otimes_S M + (y - f)F \subset$

$S' \otimes_S F$. (Note that if M is a graded module and f is homogeneous of degree d, we should take y to have degree d to get a graded module M'.) Show that $(M' : y) = (y - f)(S' \otimes_S F) + S'(M : f)$, and that $(M : f) = (M' : y) \cap F$. Deduce corresponding formulas for $(M : f^\infty)$.

d. (Reduction to the homogeneous case.) Independent of the reductions above, the computation of $(M : J)$ and $(M : J^\infty)$ can be reduced to the homogeneous case as follows: Suppose $M \subset F$ is an arbitrary submodule, let x_0 be a new indeterminate, and let $\tilde{M} \subset S[x_0] \otimes_S F$ be an $S[x_0]$-module obtained by homogenizing with respect to x_0 any set of generators for M—that is, by regarding the generators of M as vectors of polynomials, and homogenizing each component of that vector to some common degree. Let $\tilde{J}$ be the ideal of $S[x_0]$ obtained by homogenizing any set of generators of J. ($\tilde{M}$ and $\tilde{J}$ depend on lots of choices.) Show that generators for $(M : J)$ may be obtained from any set of generators for $(\tilde{M} : \tilde{J})$ by setting x_0 to 1, and similarly for $(M : J^\infty)$.

Exercise 15.42 (Another way to compute intersections): If $I = (f_1, \ldots, f_s)$ and $J = (g_1, \ldots, g_t)$ are ideals of S, show that the kernel of the map $S^{s+t+1} \to S^2$ with matrix

$$\begin{pmatrix} 1 & f_1 & \ldots & f_s & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 & g_1 & \ldots & g_t \end{pmatrix}$$

consists of vectors whose first coordinates generate the ideal $I \cap J$. Generalize this to a construction for the intersection of two submodules of an arbitrary free module.

Exercise 15.43 (Yet another way to compute intersections): If I and J are ideals of S, define an ideal K of $S[t]$ as $K = (tI + (1-t)J)$. Show that $I \cap J = K \cap S$, reducing the problem of intersection to a problem of elimination.

The following sequence of exercises provides applications for the pullback construction described in the chapter.

Exercise 15.44 (Kernels): Prove Proposition 15.32, constructing kernels.

Exercise 15.45 (Images): Let $\varphi : G \to F$ be a map of free modules, $P = \operatorname{coker} \varphi$. Given a map of free modules $H \to F$, use the pullback construction to find a presentation of the module that is the image of H in P.

Exercise 15.46: Let

$$G \xrightarrow{\varphi} F \to P \to 0$$

be a free presentation of a module P, and let M be the image of φ, so that $P = F/M$.

a. (The annihilator of an element of P.) Given an element $\bar{e}$ of P, choose $e \in F$ mapping to $\bar{e}$. Define a map $S \to F$ by sending 1 to e. Show that the annihilator of $\bar{e}$ is the image of the map π_S in the pullback diagram

$$\begin{array}{ccc} PB & \xrightarrow{\pi_S} & S \\ \downarrow & & \downarrow \\ G & \to & F \end{array}$$

b. (Annihilators in general) To compute the annihilator of P itself one can compute the annihilator of each of a set of generators, and take the intersection; however, there is a convenient way of doing this all at once: Choosing a basis $\{e_i\}$ of F, let $\psi : S \to \operatorname{Hom}(F, F)$ be the map sending 1 to the identity map. Write $\operatorname{Hom}(F, \varphi)$ for the map $\operatorname{Hom}(F, G) \to \operatorname{Hom}(F, F)$ induced by φ. Show that the annihilator of P is the image of the map π_S in the pullback diagram

$$\begin{array}{ccc} PB & \xrightarrow{\pi_S} & S \\ \downarrow & & \downarrow \psi \\ \operatorname{Hom}(F,G) & \xrightarrow[\operatorname{Hom}(F,\varphi)]{} & \operatorname{Hom}(F,F) \end{array}$$

c. (Quotient by an element) If $g \in S$, show that the submodule $(M : g) \subset F$ is the image of π_F in the pullback diagram

$$\begin{array}{ccc} PB & \xrightarrow{\pi_F} & F \\ \downarrow & & \downarrow \psi \\ G & \xrightarrow[\varphi]{} & F \end{array}$$

where $\psi : F \to F$ is multiplication by g.

d. (Quotients in general) If $J \subset S$ is an ideal with t generators $g_1, \ldots, g_t$, we could compute $(M : J)$ from the formula $(M : J) = \cap(M : g_i)$, but we can do it all at once as follows: Define a map $\alpha : S \to S^t$ sending $1 \in S$ to the column vector with ith entry g_i. Let $F \otimes \alpha : F = F \otimes S \to F \otimes S^t$ be the tensor product of the identity map and α, and let $\varphi \otimes S^t$ be the tensor product of φ with the identity map of S^t. Show that the submodule $(M : J)$ is image of π_F in the pullback diagram

$$\begin{array}{ccc} PB & \xrightarrow{\pi_F} & F \\ \downarrow & & \downarrow F\otimes\alpha \\ G \otimes S^t & \xrightarrow[\varphi\otimes S^t]{} & F \otimes S^t \end{array}$$

15.12 Appendix: Some Computer Algebra Projects

Several current computer algebra systems allow the computation of Gröbner bases. Unfortunately, as of this writing the general-purpose systems such as Macsyma, Maple, Mathematica, and Axiom do not have the flexibility in their algorithms or simply do not run fast enough to make experimentation of the sort suggested below very attractive. At least two systems that were designed primarily for Gröbner basis computation are generally available (and for free!): **CoCoA** (Computations in Commutative Algebra) by Alessandro Giovini and Gianfranco Niesi, of the Department of Mathematics, University of Genova, Italy, and **Macaulay**, by Dave Bayer and Michael Stillman, Department of Mathematics, Columbia University and Cornell University, respectively.

Macaulay is available free from its authors for many machines, including the Macintosh, IBM-PC, Sun, Vax, and others. It can be obtained from a public account on a machine at Harvard University. For experts the following instructions should suffice:

ftp math.harvard.edu, login ftp, password any, cd Macaulay.

Documentation, the source code of the program, a "make" file for compiling it on unix machines, and precompiled versions for the Macintosh and IBM compatible machines, are in compressed files in this directory. A file called "readme" describes how to make use of them.

CoCoA is relatively easy to use and is well suited for experimentation with Gröbner bases, although it lacks many of the facilities that the more mature system Macaulay has developed for handling problems from commutative algebra and algebraic geometry. On the other hand, certain design decisions taken to make Macaulay efficient may look odd to the beginner: Macaulay only computes Gröbner bases of homogeneous ideals, and works exclusively over finite fields $\mathbf{Z}/p$, for various p. In any case, I have mainly had experience with Macaulay and this appendix is slanted toward its use.

Macaulay is partially "responsible" for quite a number of published theorems, in the sense that people have been able to look at examples that have lead them to guess at results, or to reassure themselves of the truth of results, which they otherwise would not have proved. I have tried to reproduce the spirit—and in some cases the topics—of some of these investigations at a suitable level below. I am certain that there are still new phenomena to be discovered in each of these realms; perhaps the student will hit on something genuinely original. With each project I have listed the names of some Macaulay commands and scripts that I would find useful if I were doing the project. (The reader can tell the difference because scripts are written beginning with the character <, while commands do not have this prefix.) If the user types

<scriptname

or

commandname

then Macaulay should provide a help message on the script or command referred to, from which the action and the correct syntax can be inferred. Of course given the rate of development of computer algebra, these suggestions are not likely to be valid for terribly long. For all the projects I would use the scripts <ring and <ideal, which make defining objects somewhat more convenient.

Project 1. Zero-Dimensional Gorenstein Ideals

Compute some ideals of the form $I = ((x_1^s, \ldots, x_r^s) : p)$, where p is a homogeneous polynomial. It's easy to do by hand the case $r = 1$ and the case p a monomial, $r =$ anything. Try the case $r = 2$ with more complicated p on the machine. (In **Macaulay**, use the "quotient" command.) How many generators does I require? Next try $r = 3$, various p. Here there is a greater range in the possible numbers of generators. Is there any restriction? Make a conjecture! How about with $r = 4$? One way to get polynomials p to try is to take random ones (made with <random_mat, for example). The answers you get in this case should depend only on r, s, and $\deg p$. What's the pattern here? Of course more possibilities will be visible if you choose very special polynomials p.

It is also interesting to use res to resolve these ideals I. Their resolutions have a certain unusual property, visible (in **Macaulay**) through the command "betti". Can you spot it? What are the possible sequences of betti numbers in the cases $r = 2, 3, 4$? Any conjectures?

There is also something funny about the Hilbert function. (In Macaulay, use hilb and <hilb_fcn.)

For your information, the ideals I that can be obtained as above are exactly what are usually called "0-dimensional, homogeneous Gorenstein ideals." See Chapter 21.

Reference: This is actually the first project that involved me personally with computer algebra. David Buchsbaum and I were interested in Gorenstein ideals around 1971–72. Ray Zibman, then an undergraduate at Brandeis, programmed the PDP 10 computer in Lisp to find the ideals I (this can be done without Gröbner bases, since in this problem all the rings involved are finite-dimensional over k). We also made a number of hand computations of the syzygies of these ideals and found a regularity in the case $r = 3$ that may not be so apparent without a good deal of study of the matrices in the resolution. You can find the results inspired by our computations in Buchsbaum and Eisenbud [1977].

Project 2. Factoring Out a General Element from an sth Syzygy

Let R be a ring. One way for an R-module $P = \operatorname{coker} \pi_P : G_P \to F_P$ to be an sth syzygy—that is, for it to be the kernel at the sth step of a free resolution—is the following: Let

$$\mathcal{H} \qquad \ldots \to H_3 \to H_2 \to F_P{}^* \to G_P{}^*$$

be the free resolution of $Q = \operatorname{coker} \pi_P{}^*$, and dualize $\mathcal{H}$ to get a complex

$$\mathcal{H}^* : \qquad G_P \to F_P \to H_2{}^* \to \ldots \to H_t{}^* \to 0 \to \ldots$$

The homology of $\mathcal{H}^*$ (kernel of one map modulo the image of the one before) at the module $H_i{}^*$ is called $\operatorname{Ext}^i_R(Q, R)$ (we rename G_P and F_P as $H_0{}^*$ and $H_1{}^*$ respectively to make this true for $i = 0$ and $i = 1$ as well). If $s \geq 1$ and $\operatorname{Ext}^i_R(Q, R) = 0$ for $i = 1, \ldots, s$, then the complex

$$0 \to m \to H_2{}^* \to \ldots \to H_{s+1}{}^*$$

is exact, so M is the sth syzygy module of coker $H_s{}^* \to H_{s+1}{}^*$. Let us say that M is a "standard s^{th} syzygy" in this case. Is every s^{th} syzygy a standard s^{th} syzygy? If so, this gives a test for whether a module is an sth syzygy; otherwise, we have defined a new notion. Try some examples to get a feel for what might be true.

Working in the case $R = S$, the polynomial ring, take a (standard) s^{th} syzygy and kill a random element. For what t is the result a (standard) t^{th} syzygy? What if you start with a free module? Perhaps the simplest case is when $P = S^t/Sf$, where S is the column vector with entries $f_1, \ldots, f_t$. Can you tell whether P is an sth syzygy from some property of the ideal generated by the f_i?

The situation is relatively simple if, as above, we work over S. Completely new phenomena—which no one understands as of this writing—arise if we replace the polynomial ring S by a factor ring, say $S/(g_1, \ldots, g_u)$. Even the case $u = 1$ is challenging—see project 3 below—but the general case seems still more baffling.

Reference: The phenomena that the reader is most likely to discover here were first noticed and exploited by W. Bruns [1976]. See for example Evans-Griffith [1985] for a general treatment of related matters.

Project 3. Resolutions over Hypersurfaces

Find some modules over $k[x, y]/(y^2)$ (For example, take any $k[x, y]$-module M and factor out y^2 times it.) Take note of whether or not y^2 was a nonzerodivisor on M; you could test for this with the script <nzd.) Resolve over $k[x, y]$, and over $k[x, y]/(y^2)$. (Use fetch; explicit length of res, as in

res I. Ires n; betti. Keep $n \leq 15$ or so.) Can you make a conjecture about the resolutions? Can you prove it? How about replacing y^2 by an arbitrary polynomial $p(x, y)$? How about in n variables?

Reference: Eisenbud [1980].

One source of examples that will probably always be interesting is the family of "rational curves" of degree d in $\mathbf{P}^r$. From an algebraic point of view rational curves are subrings

$$R = k[f_0(s,t), f_1(s,t), \ldots, f_r(s,t)] \subset k[s,t]$$

of a polynomial ring in two variables generated by $r+1$ independent forms of degree d. (Use <subring and, for the special case where all the f_i are monomials, <monomial_curve to construct these conveniently.) The **defining ideal** of the curve is the kernel I of the map

$$k[x_0, \ldots, x_r] \to k[s,t]; x_i \mapsto f_i.$$

The next three projects explore various aspects of these examples.

Project 4. Rational Curves of Degree $r+1$ in $\mathbf{P}^r$

Consider the case of the subring R generated by 4 polynomials $f_0(s,t), \ldots, f_3(s,t)$ of degree 4 in $k[s,t]$. For various choices, compute the defining ideal I in $k[x_0, \ldots, x_3]$ and its free resolution. How do the betti numbers depend on the polynomials chosen? How many types are there? (Try the monomial examples first, then something just a little more general. Note that the result depends only on the vector space spanned by the f_i, not on the f_i themselves.) What about the situation of r forms of degree r? Note that we are excluding just one form; that is, the space of r forms of degree r is the kernel of a linear form on the space of all forms of degree r. (In Macaulay, use the command diff.) One way to write down such a linear form is as a differential operator of order r with constant coefficients—that is, essentially, as a single polynomial of degree r. This point of view may make the results more intelligible by giving natural invariants of such codimension-1 subspaces—for example, you might distinguish a single polynomial g of degree d by the smallest number t such that g is expressible as the sum of t d^{th} powers of linear forms, or is in the closure of the set of polynomials that are expressible this way. (If you like this point of view, you might want to look up "Catalecticants" in the older book on invariant theory by Grace and Young [1903].)

Project 5. Regularity of Rational Curves

If M is a graded S-module, then we define the regularity of M (in the sense of Castelnuovo) from the minimal free resolution of M

$$\ldots \to F_s \to \ldots \to F_0 \to M \to 0$$

to be the least integer ρ such that for each s, all the free generators of F_s lie in degree $\leq s + \rho$ (see Chapter 20). In Macaulay, the command res always computes minimal free resolutions after the first step; use nres to make the first step minimal too. Thus, in Macaulay the regularity is the number of rows in the diagram produced by the command betti. The regularity of M is an important measure of how hard it will be to compute a free resolution of M.

What is the possible regularity of S/I if I is the defining ideal of a rational curve? Try monomial curves (where all the f_i are monomials) first. What range of values can you get? Another interesting invariant to study in these cases is the last betti number of the curve. One way (of many) to produce interesting families of monomial curves is to fix a pattern of exponents—that is, an increasing sequence of numbers $b_1, \ldots, b_r$—and try something like

$$1, t^{a+b_1}, t^{a+b_2}, \ldots, t^{a+b_r}$$

for varying a.

Reference: Gruson, Lazarsfeld, and Peskine [1983].
(Helpful Macaulay scripts: <regularity, <res, <random_mat, <monomial_curve)

Project 6. Some Monomial Curve Singularities

Let

$$f_i = s^{d_i} t^{e_i} \quad \text{with } d_i + e_i = d, \qquad 0 \leq e_0 \leq \ldots \leq e_r \leq d,$$

and consider the corresponding rational curve. Show that factoring out $s^{d_r} t^{e_0}$ from each of the f_i will not change the defining ideal of the curve, so we may assume $e_0 = 0, e_r = d$.

Dehomogenize the defining ideal $I \subset k[x_0, \ldots, x_r]$ of the curve by setting $x_0 = 1$ (this will be the defining ideal of the subring $k[t^{e_1}, \ldots, t^{e_r}] \subset k[t]$). Compute the associated graded ring of the curve with respect to the maximal ideal. (In Macaulay use the script <l_tangentcone.)

What are the possible lengths of the minimal free resolution of this graded ring? Can you find any families of examples where the length is $r - 1$? Can you find any where the betti numbers (ranks of the free modules in the resolution) are symmetric around the middle? Try patterns of exponents, as described in Project 5.

Project 7. Some Interesting Prime Ideals

For each $0 \leq u \leq r$, consider the prime ideal $I_{u,r}$ that is the kernel of the ring homomorphism

$$k[x_0, \ldots, x_r] \to k[s, t, z_0, \ldots, z_u]; \qquad x_i \mapsto F_i,$$

where the elements F_i are obtained as homogenizations of degree $r+1$

$$F_i = s^n \varphi_i(t/s, z_0/s, \ldots, z_u/s)$$

of the entries φ_i of the product shown in Figure 15.8,

$$(z_0, z_1, \ldots, z_u) \begin{pmatrix} 1 & t & t^2 & t^3 & & \ldots & t^r \\ 0 & 1 & 2t & 3t^2 & & \ldots & rt^{r-1} \\ 0 & 0 & 2 & 6x & & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\ 0 & \ldots & 0 & & u! & \ldots & r!/(r-u)!t^{r-u} \end{pmatrix} = (\varphi_0, \varphi_1, \ldots, \varphi_r)$$

FIGURE 15.8.

where the rows of the large matrix are obtained by successively differentiating the entries of the first row with respect to t. Thus, for example, for $u = 0$ the F_i are $z_0 s^{r-1} t, \ldots, z_0 t^r$, while for $u = 1$ we get

$$z_0 s^r, \quad z_0 s^{r-1} t + z_1 s^r, \quad z_0 s^{r-2} t^2 + 2 z_1 s^{r-1} t, \ldots .$$

What degree elements do you think it takes to generate $I_{u,r}$ if r is rather larger than u? For $u = 1$ the resolution of $I_{u,r}$ has a particularly interesting property; can you find it? Can you see any interesting properties of the resolution for other values of s? In general, how long do you think the resolution will be? (You will probably have to guess at these answers from rather small values of r, u—say $r \leq 9$ or 10, $u = 0$, 1,2 and perhaps a little more.) Suppose you take the ideal generated by just the quadratic forms in $I_{u,r}$ (respectively, forms of degree $\leq d$ for some d). Do you get anything interesting?

(Use the commands power, diff, concat, to form the big matrix; mult to form the row of φ_i; homog (applied to the transposed vector) to homogenize it. Use <subring to compute $I_{u,r}$. Note that one must then use std or nres to get a minimal set of generators for the ideal.)

These ideals arise in geometry as follows. The vector $(1, t, t^2, \ldots, t^r)$ that is the first row of the preceding matrix may be thought of as parametrizing a curve C in $\mathbf{P}^r$ whose closure is called a **rational normal curve**. Thus $I_{0,r}$ is the ideal of the rational normal curve in $\mathbf{P}^r$.

The second row of the matrix is obtained by differentiating the first row with respect to x; thus a linear combination z_0(first row) + z_1(second row) represents a point on a tangent line to this curve, and $I_{1,r}$ is the ideal of the **tangent developable surface to the rational normal curve** (that is, the surface consisting of the union of the tangent lines to the curve). Similarly, for arbitrary u, the linear combination of the $u+1$ rows represents a point on an **osculating u-plane** to the curve; thus $I_{u,r}$ is the ideal of the union of the osculating u planes. These ideals have been much studied for

$u = 0$ (easy) and for $u = 1$ (the "generic Green's conjecture" is a guess at the form of the free resolution of $I_{1,r}$, see Eisenbud [1992] for an exposition). I think there are not even conjectures for $u > 1$; perhaps the reader will make some interesting ones!

16

Modules of Differentials

In this chapter we shall study objects that play roles in commutative algebra and algebraic geometry analogous to those of the tangent and cotangent bundles in the geometry of manifolds.

Definition. *If S is a ring and M is an S-module, then a map (of abelian groups) $d : S \to M$ is a* ***derivation*** *if it satisfies the* ***Leibniz rule***

$$d(fg) = fdg + gdf \quad \text{for } f, g \in S.$$

If S is an R-algebra, then we say that d is ***R-linear*** *if it is a map of R-modules. The set $\mathrm{Der}_R(S, M)$ of all R-linear derivations $S \to M$ is naturally an S-module, with multiplication defined by*

$$bd : f \mapsto b(d(f)) \in M.$$

For a familiar example let $S = k[x, y]$ be a polynomial ring in two variables. The partial derivative $\partial/\partial x$ is a derivation from S to itself. This derivation is $k[y]$-linear, and in fact the module $\mathrm{Der}_{k[y]}(k[x, y], k[x, y])$ is a free $k[x, y]$-module of rank 1, generated by $\partial/\partial x$ (see Proposition 16.1).

It is most interesting in practice to choose $M = S$ and study $\mathrm{Der}_R(S, S)$. One source of this interest is the case where S is coordinate ring of an affine variety X defined over a field R. As we shall see in more detail below, $\mathrm{Der}_R(S, S)$ is then the set of algebraic tangent vector fields on X.

For any derivation d we have $d(1) = 0$, as one sees by subtracting $d(1)$ from both sides of the equation

$$d(1 \cdot 1) = 1d(1) + 1d(1).$$

It follows that d is R-linear iff $da = 0$ for every $a \in R$: Namely, if d is R-linear then

$$da = d(a \cdot 1) = ad1 = 0,$$

and the converse is immediate from the Leibniz rule.

A dual view of derivations may be had by means of the following extremely important device:

Definition. *If S is an R-algebra, then the* ***module of Kähler differentials*** *of S over R, written $\Omega_{S/R}$, is the S-module generated by the set $\{d(f)|f \in S\}$ subject to the relations*

$$d(bb') = bd(b') + b'd(b) \quad \textit{(Leibniz)}$$
$$d(ab + a'b') = ad(b) + a'd(b') \quad \textit{(R-linearity)}$$

for all $a, a' \in R$, and $b, b' \in S$. We often write df instead of $d(f)$. The map $d : S \to \Omega_{S/R}$ defined by $d : f \mapsto df$ is an R-linear derivation, called the ***universal R-linear derivation.***

The map d has, from the definition, the following ***universal property*** (which determines it and $\Omega_{S/R}$ uniquely): Given any S-module M and R-linear derivation $e : S \to M$, there is a unique S-linear homomorphism $e' : \Omega_{S/R} \to M$ such that $e = e'd$, as in Figure 16.1.

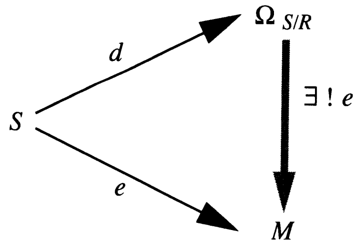


FIGURE 16.1.

Indeed, e' is defined by the formula $e'(df) = ef$. This formula defines a homomorphism because the relations among the df are also satisfied by the ef. Asserting this universal property is the same as asserting that

$$\mathrm{Der}_R(S, M) \cong \mathrm{Hom}_S(\Omega_{S/R}, M)$$

naturally, as functors of M. In this sense the construction of $\Omega_{S/R}$ "linearizes" the construction of derivations. Since the formula above allows us to compute $\mathrm{Der}_R(S, M)$ in terms of $\Omega_{S/R}$, we shall concentrate mostly on $\Omega_{S/R}$ in what follows.

If S is generated as an R-algebra by elements f_i, then $\Omega_{S/R}$ is generated as an S-module by the elements df_i: For if $g = p(f_1, \ldots, f_r)$ is a polynomial in the f_i with coefficients in R, then repeated use of the Leibniz rule allows us to express dg as an S-linear combination of the df_i. In particular, $\Omega_{S/R}$

is finitely generated as an S-module whenever S is finitely generated as an R-algebra, despite the very nonfinite nature of the definition of $\Omega_{S/R}$.

So as to have at least one example in hand before going further, we investigate the case of a polynomial ring:

Proposition 16.1. *If $S = R[x_1, \ldots, x_r]$, the polynomial ring in r variables, then $\Omega_{S/R} = \oplus_{i=1}^{r} S dx_i$, the free module on the dx_i.*

Note that we may regard S as the tensor product, over R, of the algebras $R[x_i]$, and that $\Omega_{S/R}$ is the corresponding direct sum. We shall see later that, in general, differentials make direct sums out of tensor products; this is another way in which forming differentials can linearize a problem.

Proof. Since S is generated as an R-algebra by the x_i, $\Omega_{S/R}$ is generated as an S-module by the dx_i and there is an epimorphism $S^r \to \Omega_{S/R}$ taking the ith basis vector to dx_i.

On the other hand, the partial derivative $\partial/\partial x_i$ is an R-linear derivation from S to S, and thus induces an S-module map $\partial_i : \Omega_{S/R} \to S$ carrying dx_i to 1 and all the other x_j to 0. Putting these maps together we get the inverse map

$$\Omega_{S/R} \xrightarrow{\begin{pmatrix} \partial_1 \\ \vdots \\ \partial_r \end{pmatrix}} S^r. \qquad \square$$

The association of an R-algebra S (or equivalently a map $R \to S$ of rings) to the S-module $\Omega_{S/R}$ and the derivation $d : S \to \Omega_{S/R}$ (as in Figure 16.2)

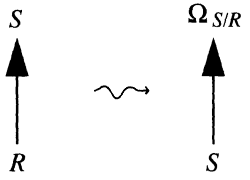
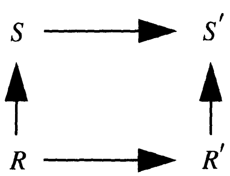


FIGURE 16.2.

is a functor in the following sense: Given a commutative diagram



of rings, which we may regard as a morphism of pairs $\varphi : (R, S) \to (R', S')$, we get an induced "morphism"

$$\begin{array}{ccc} \Omega_{S/R} & \longrightarrow & \Omega_{S'/R'} \\ d \uparrow & & \uparrow d \\ S & \longrightarrow & S' \end{array}$$

where the bottom horizontal map is the given morphism of R-algebras and the upper horizontal map is a morphism of S-modules (the module $\Omega_{S'/R'}$ being considered an S-module by means of the bottom map). The necessary S-linear map $\Omega_{S/R} \to \Omega_{S'/R'}$ is obtained from the universal property of $\Omega_{S/R}$, applied to the R-linear derivation $S \to \Omega_{S'/R'}$ that is the composition of the given map $S \to S'$ with the universal R'-linear derivation $S' \to \Omega_{S'/R'}$. (This is more complicated to say than to prove!) In applications the map $R \to R'$ is often the identity, and we think of φ as a homomorphism from S to S'. We shall sometimes replace the S-linear map $\Omega_{S/R} \to \Omega_{S'/R'}$ with the equivalent data of the S'-linear map $S' \otimes_S \Omega_{S/R} \to \Omega_{S'/R'}$. Also, we sometimes suppress mention of the universal derivation d, and speak of $\Omega_{S/R}$ itself as a functor.

For reasons we shall soon describe, $\Omega_{S/R}$ is often called the **relative cotangent functor.** It is right-exact in the same sense that the zeroeth relative homology functor is a right-exact functor of pairs of spaces in topology.

Proposition 16.2 (Relative Cotangent Sequence). *If $R \to S \to T$ are maps of rings, then there is a right-exact sequence of T-modules*

$$T \otimes_S \Omega_{S/R} \to \Omega_{T/R} \to \Omega_{T/S} \to 0$$

where the right-hand map takes dc to dc and the left-hand map takes $c \otimes db$ to cdb.

Proof. The generators for $\Omega_{T/S}$ (as a T-module) are the same as the generators of $\Omega_{T/R}$, but there are extra relations, of the form $db = 0$ for $b \in S$. These relations are precisely the images of the generators $1 \otimes db$ of the module on the left.

By analogy with the relative homology functor, one might hope for a homology theory that would give a long exact sequence of "higher cotangent functors" continuing this relative cotangent sequence to the left. The desired theory exists, and the functors are called "André-Quillen homology modules," or "higher cotangent functors." The first few, at least, have elementary and important applications to deformation theory. See Quillen [1970] or André [1974] for the general construction.

Returning to the setting of Proposition 16.2, suppose that $S \to T$ is an epimorphism. In this case $\Omega_{T/S} = 0$ (in general, $dc = 0$ whenever c is in the image of S, by S-linearity), but the next functor to the left in the exact sequence (the first of the André-Quillen homology functors, in this case) is

easy to describe, so we get another very useful exact sequence, called the **conormal sequence** (the name comes from the fact that I/I^2 is called the **conormal module** of T/S).

Proposition 16.3 (Conormal Sequence). *If $\pi : S \to T$ is an epimorphism of R-algebras, with kernel I, then there is an exact sequence of T-modules*

$$I/I^2 \xrightarrow{d} T \otimes_S \Omega_{S/R} \xrightarrow{D\pi} \Omega_{T/R} \to 0$$

where the right-hand map is given by $D\pi : c \otimes db \mapsto cdb$ and the left-hand map takes the class of f to $1 \otimes df$.

Proof. Consider the map $d : I \to \Omega_{S/R}$ that is the restriction of the universal derivation $S \to \Omega_{S/R}$. If $b \in S$ and $c \in I$, then the Leibniz formula $d(bc) = bd(c) + cd(b)$ shows that d induces an S-linear map $I \to (\Omega_{S/R})/(I\Omega_{S/R}) = T \otimes_S \Omega_{S/R}$. Taking $b \in I$ as well, the same formula shows that I^2 goes to 0 in $T \otimes_S \Omega_{S/R}$, so we get a map of T-modules

$$d : I/I^2 \to T \otimes_S \Omega_{S/R}$$

as described in the statement.

To show that the cokernel of this map is given by $D\pi$, we consider how to describe $T \otimes_S \Omega_{S/R}$ by generators and relations: From the definition of $\Omega_{S/R}$, and the right-exactness of tensor products, we see that $T \otimes_S \Omega_{S/R}$ is generated as a T-module by the elements db for $b \in S$, subject to the relations of R-linearity and the Leibniz rule. This is the same as the description by generators and relations of $\Omega_{T/R}$, except that in $\Omega_{T/R}$ the elements df for $f \in I$ are replaced by $d0$, which is of course 0 since $0 \in R$. Thus $\Omega_{T/R}$ is the cokernel of $d : I/I^2 \to T \otimes_S \Omega_{S/R}$ as claimed. □

The question of when the left-hand map of the conormal sequence is an injection is subtle, but a fundamental computation allows us to say when it is a split injection; see Proposition 16.12. It is also known to be injective when I is a radical ideal generated by a regular sequence in a polynomial ring, and somewhat more generally. See Exercise 16.17.

16.1 Computation of Differentials

Nearly all explicit computations of modules of differentials use these ideas in a simple way that we may formalize as follows: If S is a finitely generated R-algebra, say $S = R[x_1, \ldots, x_r]/I$, and if $I = (f_1, \ldots, f_s)$, then Proposition 16.1 shows that $S \otimes_R \Omega_{R[x_1,\ldots,x_r]/R} = \oplus_i S dx_i$ is a free S-module on generators dx_i, and by the conormal sequence

$$\Omega_{S/R} = \operatorname{coker}(d : I/I^2 \to \oplus_i S dx_i).$$

Writing I/I^2 as a homomorphic image of a free S-module with generators e_i going to the classes of the f_i, the composition

$$\mathfrak{J} : \oplus Se_i \twoheadrightarrow I/I^2 \to \oplus_i Sdx_i$$

is a map of free S-modules that is represented by the Jacobian matrix of the f_j with respect to the x_i: That is, the (i, j) entry of $\mathfrak{J}$ is $\partial f_j/\partial x_i$. In short, $\Omega_{S/R}$ is the cokernel of the Jacobian matrix $\mathfrak{J} = (\partial f_j/\partial x_i)$, regarded as a map of free S-modules. For example, if $S = R[x]/f(x)$, then

$$\Omega_{S/R} = Sdx/df = Sdx/(S \cdot f'(x)dx) \cong S/(f'(x)).$$

This idea can sometimes be applied even to algebras that are not finitely generated over R. For example, if we are interested in a localization of an algebra of finite type, say $S' = S[U^{-1}]$ for some multiplicatively closed set U, then applying Proposition 16.9 we see that $\Omega_{S'/R}$ is the cokernel of the same Jacobian matrix, now thought of as a map of free S'-modules.

For an explicit example, consider the ring $S = R[x, y, t]/(y^2 - x^2(t^2 - x))$. In this case $\mathfrak{J}$ is the 3×1 matrix

$$\mathfrak{J} = \begin{pmatrix} 3x^2 - 2xt^2 \\ 2y \\ -2x^2t \end{pmatrix}$$

and the computation above shows that $\Omega_{S/R}$ is the free S-module on the generators dx, dy, and dt modulo the single relation

$$(3x^2 - 2xt^2)dx + (2y)dy - (2x^2t)dt = 0.$$

We shall return to this example in the section on the Jacobian Criterion.

16.2 Differentials and the Cotangent Bundle

We have already said that modules of differentials have something to do with tangent bundles, and we shall now describe the connection more precisely. (The reader who is not familiar with the notion of the tangent bundle of a smooth manifold may skip this discussion; alternately, the little information that we shall need about this basic notion may be rapidly acquired from many sources, among them Hirsch [1976].)

Briefly, the connection is this: If Y is an affine algebraic variety over a field k with coordinate ring S, then $\Omega_{S/k}$ plays the role of the cotangent bundle of Y. More generally, if $Y \to X$ is a morphism of affine varieties corresponding to a map $R \to S$ of coordinate rings, then $\Omega_{S/R}$ plays the role of the relative cotangent bundle of the map.

Here is a more detailed description: For every smooth (differentiable is enough) manifold X (over $k = \mathbf{R}$, say) there is a vector bundle on X, called the tangent bundle of X and written T_X, whose fiber over a point $x \in X$

is the tangent space $T_{X,x}$ of X at x. If $\varphi : X \to Y$ is a differentiable map of smooth manifolds, then for every $x \in X$ the derivative of φ is a map T_φ from $T_{X,x}$ to the tangent space of Y at $\varphi(x)$, that is, to $T_{Y,\varphi(x)}$. These derivatives fit together into a map of vector bundles on X

$$T\varphi : T_X \to \varphi^* T_Y,$$

where $\varphi^* T_Y$ is the tangent bundle to Y "pulled back" along φ (the pullback may be defined as the fiber product $X \times_Y T_Y$, so that the fiber over x of $\varphi^* T_Y$ is $T_{Y,f(x)}$).

To connect these constructions with our previous constructions, let S' be the ring of smooth functions on X. For any $f \in S'$, thought of as a mapping to the line $\mathbf{R}$, the derivative $Tf : T_X \to \varphi^* T_{\mathbf{R}} = X \times \mathbf{R}$ is a linear functional on each fiber $T_{X,x}$ that varies smoothly with x. Thus Tf may be considered to be a global section of the dual T_X^* of the tangent bundle, which is called the cotangent bundle of X. Of course, if g is another function, then $T(fg) = fTg + gTf$, so we may think of T as a derivation of the ring S' of smooth functions on X to the S'-module Ω' of global sections of the cotangent bundle of X.

From the universal property of the module of Kähler differentials it follows that there is an S'-module homomorphism $\alpha : \Omega_{S'/k} \to \Omega'$ carrying the universal derivation d to the derivation T just constructed. This is usually not an isomorphism, essentially because we have not taken topology into account in defining $\Omega_{S'/k}$, but if X is actually a real affine variety and S is its coordinate ring, then it can be shown that $\Omega_{S/k}$ is the algebraic object precisely analogous to Ω' in the sense that $\Omega' = \Omega_{S/k} \otimes_S S'$. Since the bundle T_X^* and the module Ω' are equivalent objects, we see that the algebraic module of differentials $\Omega_{S/k}$ is a good stand-in for the cotangent bundle. Similarly, its dual $\mathrm{Der}_R(S, S)$ is a satisfactory replacement for the tangent bundle.

If $\varphi : X \to Y$ is a map of manifolds, we write $\varphi^\#$ for the homomorphism from the ring of smooth functions on Y to the ring of smooth functions on X that is given by composition with φ. The dual of the map $T\varphi : T_X \to \varphi^* T_Y$ is a map $\varphi^*(T_Y^*) \to T_X^*$. This is the map that is analogous to the map $D(\varphi^\#)$.

The construction we have made has a somewhat more general analogue that is worth keeping in mind. If $\varphi : X \to Y$ is a map of manifolds whose derivative $T\varphi$ is everywhere surjective (such a map is said to be "submersive"), then the kernel of $T\varphi : T_X \to \varphi^* T_Y$ is again a bundle on X, called the **relative tangent bundle,** $T_{X/Y}$, and its dual is the **relative cotangent bundle** $T_{X/Y}^*$. The bundle $T_{X/Y}$ may be thought of as the bundle of tangent vectors that are tangent to fibers of φ. Given any function f on X, Tf restricts to a linear functional on $T_{X/Y}$. If f is the composition of φ with a smooth function g on Y, then f will be constant on the fibers, and Tf will induce the 0 functional on $T_{X/Y}$; that is, $Tf = 0$ in $T_{X/Y}^*$.

From this we deduce that T, as a map from functions on X to sections of $T^*_{X/Y}$, is an R'-linear derivation, where R' is the ring of smooth functions on X. If X and Y are affine varieties with coordinate rings S and R, respectively, so that S is an R-algebra, then the module $\Omega_{S/R}$ with its universal R-linear derivation from S is the algebraic version of this relative cotangent construction.

In the geometric case of nonsingular manifolds that we have sketched, the notions of tangent and cotangent bundles have the same content; each bundle is obtained from the other by dualization. Rather than looking at the module of differentials, one might just as well look at the module of tangent vector fields, which may be thought of as derivations. But when dealing with algebraic varieties with singularities, this equivalence no longer holds: Derivations may be derived from differential forms, but not conversely. That is why we usually work with $\Omega_{S/R}$ rather than with its dual $\mathrm{Der}_R(S,S)$.

The geometric meaning of Proposition 16.2 is as follows: We defined the relative tangent bundle coming from a submersive map $\varphi : X \to Y$ of manifolds as the kernel of the map $T_X \to \varphi^* T_Y$, and correspondingly the relative cotangent bundle is $T^*_{X/Y} = T^*_X/\mathrm{image}(\varphi^* T^*_Y)$. If $Z \to X \to Y$ are two submersive maps of manifolds, then simply from this definition it follows that there is an exact sequence

$$\varphi^*(T^*_{X/Y}) \to T^*_{Z/Y} \to T^*_{Z/X} \to 0,$$

which corresponds to our relative cotangent sequence.

The name conormal sequence in Proposition 16.3 also comes from the geometric case. Given a submanifold X of a manifold Y, we may restrict the tangent bundle of Y to X and get a vector bundle on X that contains the tangent bundle to $X, T_X \hookrightarrow T_{Y|_X}$. The cokernel of this inclusion is called the normal bundle of X in Y, written $N_{X/Y}$. Part of its significance comes from the fact that $N_{X/Y}$ resembles a tubular neighborhood of X in Y: In the differentiable category they are isomorphic (though if X and Y are analytic manifolds, they are generally not isomorphic in the analytic category).

From the definition, we have an exact sequence of bundles

$$0 \to T_X \to T_{Y|_X} \to N_{X/Y} \to 0,$$

which might be called the "normal sequence of X in Y." Dualizing, we get the conormal sequence of X in Y:

$$0 \to N^*_{X/Y} \to T^*_{Y|_X} \to T^*_X \to 0.$$

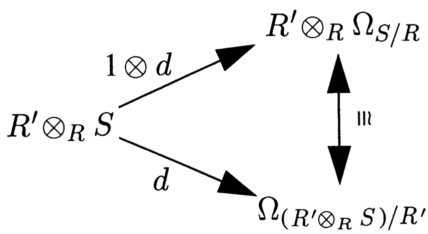
In an algebraic setting, working over a ground field $R(= \mathbf{R}$ or $\mathbf{C}$ in the case above), suppose that Y is an affine variety with coordinate ring T, and that X is a closed affine subvariety with coordinate ring $T = S/I$. Then T^*_X corresponds to $\Omega_{T/R}$ and $T^*_{Y|_X}$ corresponds to $T \otimes_R \Omega_{S/R}$, so that I/I^2 corresponds to the conormal bundle $N^*_{X/Y}$ and the sequence in Proposition 16.3

is indeed the conormal sequence. The only thing that appears different in our general algebraic setting is that the map $I/I^2 \to T \otimes_R \Omega_{S/R}$ is not always injective. But recall that, in the algebraic setting, X is an arbitrary affine subvariety, not a submanifold. It turns out that if we assume something corresponding to the assumption that X is a submanifold of Y(X locally a complete intersection in Y in the sense of Chapter 18) then I/I^2 will be locally free —that is, really corresponds to a vector bundle —and the map $I/I^2 \to T \otimes_R \Omega_{S/R}$ will be an inclusion, bringing us exactly into line with the classical geometric situation. See Exercise 16.17.

16.3 Colimits and Localization

We now list some tools that make working with modules of differentials convenient.

Proposition 16.4 (Base Change). *Formation of differentials commutes with arbitrary "base change from R"; that is, for any R-algebras R' and S there is a commutative diagram as follows.*



Proof. We use the universal properties to get vertical maps as in the diagram: First, $1 \otimes_R d : R' \otimes S \to R' \otimes \Omega_{R/S}$ is an R'-linear derivation, so there is a map $\Omega_{(R' \otimes_R S)/R'} \to R' \otimes_R \Omega_{S/R}$ sending $d(a' \otimes b)$ to $a' \otimes d(b)$. To go the other way, note that the composite map

$$S = R \otimes_R S \to R' \otimes_R S \xrightarrow{d} \Omega_{(R' \otimes_R S)/R'}$$

is an R-linear derivation, so that there is a map of S-modules $\Omega_{S/R} \to \Omega_{(R' \otimes_R S)/R'}$ sending db to $d(1 \otimes b)$. Since the target is an $R' \otimes_R S$-module, this induces an $R' \otimes_R S$-linear map

$$R' \otimes_R \Omega_{S/R} \to \Omega_{(R' \otimes_R S)/R'}$$

sending $a' \otimes db$ to $a'db = d(a' \otimes b)$, and this is the inverse of the previous map. □

Unfortunately, no such result holds for general maps $S \to S'$; but we shall see below that there is a similar formula for the case of a localization of S.

On the other hand, since $S \to \Omega_{S/R}$ has a universal property, one should expect on categorical grounds that it should "preserve colimits" in some suitable sense (see Appendix 6 for information about colimits, and Exercises 2h and 4 in Appendix 5 for another view of the following results). We give two special cases that together contain the substance of this statement, and then formulate the general categorical result as Theorem 16.8.

The colimits of greatest interest for us are the coproducts, and we explain these first. In the category of R-algebras the coproduct of a (possibly infinite) set of algebras $\{S_i\}$ is the **restricted tensor product** of the S_i (Proposition A6.7b), which we shall write as $\otimes_{R,i} S_i$ or $\otimes_R S_i$. Recall that this is the algebra generated by the symbols

$$b_1 \otimes b_2 \otimes \cdots \quad \text{with } b_i \in S_i, \quad b_i = 1 \quad \text{for all but finitely many } i,$$

modulo the relations of R-multilinearity

$$\begin{aligned} b_1 \otimes b_2 \otimes \cdots \otimes \ (ab_i + a'b_i') \otimes \cdots = \\ a(b_1 \otimes b_2 \otimes \cdots \otimes b_i \otimes \cdots) + a'(b_1 \otimes b_2 \otimes \cdots \otimes b_i' \otimes \cdots) \end{aligned}$$

for $a, a' \in R$ and $b_i, b_i' \in S_i$, with multiplication defined componentwise.

To simplify the notation, we shall exploit the commutativity of the tensor product and allow ourselves to write tensors in any order.

Proposition 16.5 (Tensor Products). *If $T = \otimes_R S_i$ is the coproduct of some R-algebras S_i, then*

$$\begin{aligned} \Omega_{T/R} &\cong \oplus_i (T \otimes_{S_i} \Omega_{S_i/R}) \\ &= \oplus_i ((\otimes_{R,j \neq i} S_j) \otimes_R \Omega_{S_i/R}) \end{aligned}$$

by an isomorphism α satisfying

$$\alpha : d(\cdots \otimes 1 \otimes 1 \otimes b_i \otimes 1 \otimes 1 \otimes \cdots) \mapsto (\ldots, 0, 0, 1 \otimes db_i, 0, 0, \ldots),$$

where $b_i \in S_i$ occurs in the ith place in each expression.

Proof. To justify the equality sign, note that

$$\begin{aligned} T \otimes_{S_i} \Omega_{S_i/R} &= (\otimes_{R,j \neq i} S_j) \otimes S_i \otimes_{S_i} \otimes \Omega_{S_i/R}. \\ &= (\otimes_{R,j \neq i} S_j) \otimes_R \Omega_{S_i/R}. \end{aligned}$$

To prove the isomorphism, let

$$\Omega := \oplus_i ((\otimes_{j \neq i} S_j) \otimes_R \Omega_{S_i/R})$$

be the direct sum, where we have written $\otimes$ for $\oplus_R$. Write $d_i : S_i \to \Omega_{S_i/R}$ for the universal derivation on S_i. Any element $c \in T$ may be written as a finite sum of terms $\otimes b_i$ with $b_i \in S_i$ and only finitely many b_i different from 1. Thus only finitely many of the maps

$$1 \otimes d_i : T = (\otimes_{j \neq i} S_j) \otimes S_i \to (\otimes_{j \neq i} S_j) \otimes_R \Omega_{S_i/R}$$

are nonzero on c, so we may define a map $e : T \to \Omega$ to be the sum $\Sigma_i 1 \otimes d_i$. Since each map $1 \otimes d_i$ is a derivation, e is too. Thus there is an induced T-module homomorphism $\alpha : \Omega_{T/R} \to \Omega$ carrying $d(\otimes_i b_i)$ to $e(\otimes_i b_i)$.

To produce the inverse of α, note that for each S_i the composite of the natural map $S_i \to T$ with the universal derivation of T is an R-linear derivation $S_i \to \Omega_{T/R}$ and thus induces an S_i-linear map $\Omega_{S_i/R} \to \Omega_{T/R}$ sending $d_i b_i$ to $d(1 \otimes b_i)$, where 1 is the identity of $(\otimes_{j \neq i} S_j)$. Since the target is a T-module, this extends to a T-linear map

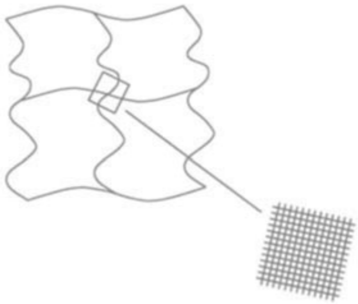
$$\beta_i : T \otimes_R \Omega_{S_i/R} \to \Omega_{T/R}$$

with

$$\beta_i : 1 \otimes d_i b_i \mapsto d(1 \otimes b_i).$$

The β_i together give a map $\Omega = \oplus_i T \otimes_R \Omega_{S_i/R} \to \Omega_{T/R}$ that is the inverse of α.

Proposition 16.5 is the geometric expression of a fundamental fact, which for simplicity we state for tensor products with just two factors. If X_1 and X_2 are manifolds, then the fiber at a point $(p_1, p_2) \in X_1 \times X_2$ of the tangent bundle to $X_1 \times X_2$ is canonically isomorphic to the direct sum $T_{X_{1,p_1}} \oplus T_{X_{2,p_2}}$, as illustrated in the figure.



Because this identification is canonical, it extends to an identification of bundles. To express it formally, we write π_1 and π_2 for the projections of $X_1 \times X_2$ to X_1 and X_2, respectively. We denote by $\pi_1^* T_{X_1}$ the pullback to $X_1 \times X_2$ of the tangent bundle to X_1, and similarly for the other projection. In these terms we have:

$$T_{X_1 \times X_2} \cong \pi_2^* T_{X_2} \oplus \pi_1^* T_{X_1}$$

In Proposition 16.5, the roles of X_1 and X_2 are played by two R-algebras S_1 and S_2, the direct product of spaces corresponds to the tensor product $T = S_1 \otimes_R S_2$, and the pullback is also given by a tensor product: If M is an S_1-module, corresponding to a bundle on X_1, then the pullback to $X_1 \times X_2$ corresponds to the $S_1 \otimes_R S_2$-module $M \otimes_R S_2$.

Proposition 16.1 exhibits a special case of Proposition 16.5. We can now go a little further in this direction.

Corollary 16.6. *If $T := S[x_1, \ldots, x_r]$ is a polynomial ring over an R-algebra S, then*

$$\Omega_{T/R} \cong T \otimes_S \Omega_{S/R} \oplus \oplus_i T dx_i.$$

Proof. Let $T' = R[x_1, \ldots, x_r]$. Writing T as $S \otimes_R T'$, we can apply the proposition to get $\Omega_{T/R} \cong T \otimes_S \Omega_{S/R} \oplus T \otimes_{T'} \Omega_{T'/R}$. By Proposition 16.1, $T \otimes_{T'} \Omega_{T'/R} = T \otimes_{T'} \oplus_i T' dx_i = \oplus_i T dx_i$. (We could also have written $T' = R[x_1] \otimes_R R[x_2] \otimes_R \cdots \otimes_R R[x_r]$, and reduced to the case $\Omega_{R[x]/R} = R[x]dx$ of Proposition 16.1.)

The other case of colimits that we must treat is that of coequalizers. The coequalizer in the category of R-algebras of a pair of maps $\psi, \psi' : S_1 \to S_2$ is the algebra $T = S_2/I$, where I is the ideal generated by all the elements $\psi(b) - \psi'(b)$ for $b \in S_1$ (Proposition A6.7c). The necessary result follows from the conormal sequence.

Corollary 16.7. *Formation of differentials preserves coequalizers in the following sense: If T is the coequalizer in the category of R-algebras of a pair of maps $\psi, \psi' : S_1 \to S_2$, then there is a right exact sequence of T-modules*

$$T \otimes_{S_1} \Omega_{S_1/R} \xrightarrow{T\otimes D\psi - T\otimes D\psi'} T \otimes_{S_2} \Omega_{S_2/R} \to \Omega_{T/R} \to 0.$$

Proof. By the conormal sequence, $\Omega_{T/R}$ is $T \otimes_{S_2} \Omega_{S_2/R}$ modulo the submodule generated by the elements $d(\psi(b) - \psi'(b))$. This submodule is the image of the map $T \otimes D\psi - T \otimes D\psi'$.

Putting these together with a basic categorical fact (Theorem A6.1), we get:

Theorem 16.8 (Colimits). *Let $\mathcal{B}$ be a diagram in the category of R-algebras. Set $\varinjlim B = T$. If F is the functor from $\mathcal{B}$ to the category of T-modules taking an object S to $T \otimes_S \Omega_{S/R}$ and a morphism $\varphi : S \to S'$ to the morphism $1 \otimes D\varphi : T \otimes_S (S \otimes_{S'} \Omega_{S'/R}) \to T \otimes_S \Omega_{S/R}$, then*

$$\Omega_{T/R} = \varinjlim F.$$

Proof. As colimits are constructed from coproducts and coequalizers, it is enough to check the proposition for each of these two types of colimits. The case of coproducts is handled by Proposition 16.5, while that of coequalizers is handled by Corollary 16.7.

As an application of these ideas we give a simple proof that modules of differentials localize well. (A direct proof is not difficult, but somewhat messier; see Exercise 16.4.)

Proposition 16.9 (Localization). *Formation of differentials commutes with localization of the upper argument; that is, if S is an R-algebra and U*

is a multiplicatively closed subset of S*, then*

$$\Omega_{S[U^{-1}]/R} \cong S[U^{-1}] \otimes_S \Omega_{S/R}$$

in such a way that $d(1/s) = -s^{-2}ds$ *for* $s \in U$*.*

Proof. Suppose first that U is the set of powers of a single element s, so that $S[U^{-1}] = S[x]/(sx-1)$. By our usual computation of modules of differentials using Corollary 16.6 and Proposition 16.3 we see that

$$\Omega_{S[U^{-1}]/R} = (S[U^{-1}]\Omega_{S/R} \oplus S[U^{-1}]dx)/(S[U^{-1}]d(sx-1)).$$

Of course $d(sx-1) = sdx + xds$, and since s is a unit in $S[U^{-1}]$, we see that $\Omega_{S[U^{-1}]/R} = S[U^{-1}]\Omega_{S/R}$, where dx is identified with $-(x/s)ds$. Thinking of x as s^{-1} this reads $d(s^{-1}) = -s^{-2}ds$, as claimed.

The general case follows from this one by a colimit argument: If $\mathcal{B}$ is the diagram of R-algebras whose objects are the localizations $S[s^{-1}]$ for $s \in U$, with maps $S[s^{-1}] \to S[(st)^{-1}]$ the natural localization maps for $s, t \in U$, then $S[U^{-1}] = \varinjlim \mathcal{B}$ (see Exercise A6.7), so by Theorem 16.8

$$\Omega_{S[U^{-1}]/R} = \varinjlim{}_{s\in U} S[U^{-1}] \otimes_{S[s^{-1}]} \Omega_{S[s^{-1}]/R}.$$

But for any S-module M,

$$M[U^{-1}] = \varinjlim{}_{s\in U} S[U^{-1}] \otimes_{S[s^{-1}]} M[s^{-1}],$$

so we are done.

The formation of the module of differentials does not commute with inverse limits in general. For example, in the case of completion $\hat{R}$ of a local ring (R, P) the module, $\Omega_{\hat{R}/R}$ is in general very large, while $\Omega_{(R/P^n)/R} = 0$ since R/P^n is a homomorphic image of R, and thus $\varprojlim \Omega_{(R/P^n)/R} = 0$. However, differentials do behave well with respect to finite direct products. Recall that the direct product of some R-algebras S_i is, as an R-module, the direct product of the R-modules S_i, with multiplication defined coordinatewise. If M_i is an S_i-module, then the product $\Pi_i M_i$ is a $\Pi_i S_i$-module, again with coordinatewise operations. In the case of a finite product, every $\Pi_i S_i$-module is obtained in this way by Exercise 2.27.

Proposition 16.10 (Direct Products). *If* $S_1, \ldots, S_n$ *are* R*-algebras and* $S = \Pi_i S_i$*, then*

$$\Omega_{S/R} = \prod_i \Omega_{S_i/R}.$$

Proof. If e_i is the idempotent of S that is the unit of S_i, and D is a derivation of S to an S-module M, then $De_i = 0$ (Exercise 16.1) so

$$D(e_i f) = e_i Df.$$

Thus D maps $S_i = e_i S$ to $M_i := e_i M$ and corresponds to a unique map $\Omega_{S_i/R} \to M_i$. It follows that $\Pi_i \Omega_{S_i/R}$ has the universal property that characterizes $\Omega_{S/R}$.

16.4 Tangent Vector Fields and Infinitesimal Morphisms

The equations defining an algebra homomorphism between two R-algebras are not linear; for this reason the set of algebra homomorphisms does not naturally form an abelian group. However, we shall show that if two algebra homomorphisms agree modulo an ideal of square 0, then they differ by a derivation. Those who know something of the theory of schemes will appreciate the geometric meaning of this: A vector field acts as a first-order infinitesimal translation. From our algebraic point of view the statement is the following:

Proposition 16.11. *Let $\varphi : S \to S'$ be a map of R-algebras, and let $\delta : S \to S'$ be a map of abelian groups. If $\delta(S)^2 = 0$ then $\varphi + \delta$ is a homomorphism of R-algebras iff δ is an R-linear derivation in the sense that $\delta(b_1 b_2) = \varphi(b_1)\delta(b_2) + \varphi(b_2)\delta(b_1)$.*

Proof. We have

$$(\varphi + \delta)(b_1 b_2) = \varphi(b_1 b_2) + \delta(b_1 b_2),$$

while

$$\begin{aligned}&(\varphi + \delta)(b_1) \cdot (\varphi + \delta)(b_2) \\ &\quad = \varphi(b_1 b_2) + \varphi(b_1)\delta(b_2) + \varphi(b_2)\delta(b_1) + \delta(b_1)\delta(b_2).\end{aligned}$$

The last term is 0, so the two expressions are equal iff $\delta(b_1 b_2) = \varphi(b_1)\delta(b_2) + \varphi(b_2)\delta(b_1)$, proving that $\varphi + \delta$ is a homomorphism of rings iff δ is a derivation. Since φ preserves R we see that $\varphi + \delta$ preserves R iff $\delta(R) = 0$. Thus $\varphi + \delta$ is a homomorphism of R-algebras iff δ is an R-linear derivation.

As a consequence, we may give a necessary and sufficient condition for the left-hand map in the conormal sequence to be a split injection.

Proposition 16.12. *If $\pi : S \to T$ is an epimorphism of R-algebras, with kernel I, then in the conormal sequence*

$$I/I^2 \xrightarrow{d} T \otimes_S \Omega_{S/R} \xrightarrow{D\pi} \Omega_{T/R} \to 0$$

the map d is a split injection iff there is a map of R-algebras $\tau : T \to S/I^2$ splitting the projection map $S/I^2 \twoheadrightarrow S/I = T$.

Proof. We first reduce to the case $I^2 = 0$. By the conormal sequence for $R \to S \twoheadrightarrow S/I^2$, the module $\Omega_{(S/I^2)/R}$ is derived from $\Omega_{S/R}$ by factoring out $I^2\Omega_{S/R}$ and $d(I^2)$. But if $a, b \in I$ then $d(ab) = ad(b) + bd(a)$ so $d(I^2) \subset I\Omega_{S/R}$. Thus $T \otimes_S \Omega_{(S/I^2)/R} = T \otimes_S \Omega_{S/R}$. Because of this we may suppose that $I^2 = 0$.

To avoid confusion with the map $d : I \to T \otimes \Omega_{S/R}$ in the conormal sequence, we shall write $d' : S \to \Omega_{S/R}$ for the universal derivation.

Suppose that d is split by a map $\sigma : T \otimes_S \Omega_{S/R} \to I$. Let $\gamma : \Omega_{S/R} = S \otimes_S \Omega_{S/R} \to T \otimes_S \Omega_{S/R}$ be the map $\pi \otimes 1$. Note that d is the restriction

of $\gamma d'$ to I. Set $\delta = \sigma\gamma d' : S \to I \subset S$; it is an R-linear derivation. By Proposition 16.11, $(1-\delta) : S \to S$ is an R-algebra homomorphism. If $b \in I$ then $\sigma db = b$, so $\delta(b) = \sigma\gamma d'(b) = \sigma d(b) = b$. Thus $(1-\delta)(I) = 0$, and $1-\delta$ induces an algebra map $\tau : T \to S$. We have $\pi\tau = \pi(1-\delta) = \pi$, so τ splits π, as claimed.

Conversely, suppose that $\tau : T \to S$ is a map of R-algebras splitting the projection π. The map $\delta := 1-\tau\pi : S \to S$ has image in the kernel of π; that is, $\delta(S) \subset I$. Since $I^2 = 0$, Proposition 16.11 shows that δ is an R-linear derivation from S to I. By the universal property of $\Omega_{S/R}$, δ corresponds to a homomorphism $\sigma' : \Omega_{S/R} \to I$. Since $I^2 = 0$ this homomorphism factors through a homomorphism $\sigma : T \otimes \Omega_{S/R} = \Omega_{S/R}/I\Omega_{S/R} \to I$.

We claim that σ is a splitting of the map $d : I \to T \otimes \Omega_{S/R}$. Indeed, if $b \in I$ then by the definition of σ and σ' we have $\sigma d(b) = \sigma' d'(b) = \delta(b)$. But the definition of δ shows that $\delta(b) = b$, and we are done.

Perhaps the most interesting case in which to apply this proposition is the case where S is a local ring and I is the maximal ideal. Recall from Chapter 7 that a **coefficient field** for any local ring $(R, \mathfrak{m})$ is a subfield of S that maps isomorphically to the residue class field $R/\mathfrak{m}$. For the notion of separability used see Appendix 1.

Corollary 16.13. *Let $(R, \mathfrak{m})$ be a local ring, and suppose that R contains a field k. Let $d : \mathfrak{m}/\mathfrak{m}^2 \to (R/\mathfrak{m}) \otimes_R \Omega_{R/k}$ be the map induced by the universal derivation $R \to \Omega_{R/k}$. The map d is a monomorphism iff there is a coefficient field for $R/\mathfrak{m}^2$ containing k. In particular, d is a monomorphism when $R/\mathfrak{m}$ is separable over k.*

Proof. As $R/\mathfrak{m}$ is a field, and d is a map of $R/\mathfrak{m}$-modules, d is an inclusion iff d is a split inclusion. Thus we may apply Proposition 16.12.

For the second statement, note that $R/\mathfrak{m}^2$ is a complete local ring. Theorem 7.8 shows that coefficient fields containing k exist if $R/\mathfrak{m}$ is separable over k. □

16.5 Differentials and Field Extensions

If $S \subset T$ are fields, then $\Omega_{T/S}$ is a vector space over T. Since $\Omega_{T/S}$ is generated by the elements $\{dx | x \in T\}$, it must have bases consisting of subsets of these elements. We say that a collection $\{x_\lambda\}_{\lambda\in\Lambda}$ of elements of T is a **differential basis** for T/S if the set $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a basis for $\Omega_{T/S}$ as a vector space over T. For example, if $T = S(\{x_\lambda\}_{\lambda\in\Lambda})$ is the field of rational functions in some set of variables $\{x_\lambda\}_{\lambda\in\Lambda}$, then it follows from Propositions 16.1, 16.9, and (for the case where Λ is infinite) Proposition 16.5, that $\Omega_{T/S}$ is free on the elements dx_λ. We shall see that in characteristic 0 the notion of differential basis coincides with the notion of transcendence

basis for T over S, while in characteristic p it coincides with the notion of a p-basis for T over S.

The notion of a p-basis is treated in detail in Appendix 1. Here we simply recall that a collection of elements $\{x_\lambda\}_{\lambda\in\Lambda} \subset T$ is a p-basis for T over S if $\{x_\lambda\}_{\lambda\in\Lambda}$ is a minimal set of generators for T as an $S * T^p$-algebra.

Theorem 16.14. *If $S \subset T$ are fields and $\{x_\lambda\}_{\lambda\in\Lambda} \subset T$ is a collection of elements, then $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a basis of $\Omega_{T/S}$ as a vector space over T iff either*

a. $\operatorname{char} S = 0$ *and $\{x_\lambda\}_{\lambda\in\Lambda}$ is a transcendence basis of T over S, or*

b. $\operatorname{char} S = p \neq 0$ *and $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a p-basis of T over S.*

In particular, if $\operatorname{char} S = 0$ *or T is a finitely generated separable field extension of S, then $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a basis of $\Omega_{T/S}$ iff the x_i form a separating transcendence basis of T over S.*

We first prove a more general result about separable algebraic field extensions.

Lemma 16.15. *Let $R \to S \subset T$ be maps of rings. If S and T are fields and T is separable and algebraic over S, then*

$$\Omega_{T/R} = T \otimes_S \Omega_{S/R}.$$

Proof. Because both the functors $\Omega_{T/S}$ and $T\otimes_S$- commute with direct limits, it suffices to prove the lemma in the case where T is a finite extension of S.

Suppose then that T is a finite separable extension of S, and choose a primitive element $\alpha \in T$. If f is the minimal polynomial of α we have $T = S[x]/(f)$. The conormal sequence for $R \to S[x] \twoheadrightarrow T$ is

$$(f)/(f^2) \xrightarrow{d} T \otimes_{S[x]} \Omega_{S[x]/R} \to \Omega_{T/R} \to 0,$$

where the left-hand map sends $f + (f^2)$ to $1 \otimes df$. Applying Corollary 16.6 we get

$$\Omega_{S[x]/R} \cong S[x] \otimes_S \Omega_{S/R} \oplus S[x]\, dx$$

so that

$$T \otimes_{S[x]} \Omega_{S[x]/R} \cong T \otimes_S \Omega_{S/R} \oplus T\, dx.$$

The component of $1 \otimes df$ in the second summand, $T\, dx$, is $f'(\alpha)\, dx$. Since T is separable over S, $f'(\alpha) \neq 0$, so $f'(\alpha)\, dx$ generates $T\, dx$. Thus $\Omega_{T/R} \cong T \otimes_S \Omega_{S/R}$.

See Exercise 16.6 for an analysis of the map $T \otimes_S \Omega_{S/R} \to \Omega_{T/R}$ in some other cases.

Proof of Theorem 16.14. We treat the two cases separately, although the proofs are parallel.

a. First suppose that $\operatorname{char} S = 0$. If $\{x_\lambda\}_{\lambda\in\Lambda}$ is a transcendence basis, then since T is algebraic over $S' = S(\{x_\lambda\}_{\lambda\in\Lambda})$, Lemma 16.15 gives $\Omega_{T/S} = T \otimes_S \Omega_{S'/S}$. Since S' is the localization (at the 0 ideal) of the polynomial ring $S[\{x_\lambda\}_{\lambda\in\Lambda}]$, Proposition 16.1 shows that $\Omega_{S'/S}$ has $\{x_\lambda\}_{\lambda\in\Lambda}$ as a basis, which gives the desired result.

Conversely, suppose that $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a basis of $\Omega_{T/S}$. Let $T' = S(\{x_\lambda\}_{\lambda\in\Lambda})$ be the subfield of T generated by $\{x_\lambda\}_{\lambda\in\Lambda}$. We first show that T is algebraic over T'. In the cotangent exact sequence

$$T \otimes_{T'} \Omega_{T'/S} \to \Omega_{T/S} \to \Omega_{T/T'} \to 0$$

the element $1 \otimes dx_\lambda \in T \otimes_{T'} \Omega_{T'/S}$ goes to $dx_\lambda \in \Omega_{T/S}$. Since these elements generate $\Omega_{T/S}$ we see that $\Omega_{T/T'} = 0$. But if $\{y_\gamma\}_{\gamma\in\Gamma}$ is a transcendence base of T over T', the argument above shows that $\{dy_\gamma\}_{\gamma\in\Gamma}$ is a basis of $\Omega_{T/T'}$. Thus Γ must be empty, and T is algebraic over T' as claimed.

Next we must show that the x_i are algebraically independent over S. If on the contrary x_1, say, were algebraically dependent on $\{x_\lambda\}_{\lambda\in\Lambda,\lambda\neq1}$, then T would be algebraic over the field $T' := S(\{x_\lambda\}_{\lambda\in\Lambda,\lambda\neq1})$. By the argument above, dx_1 would be in the submodule generated by the $\{dx_\lambda\}_{\lambda\in\Lambda,\lambda\neq1}$, contradicting the hypothesis that the dx_λ are linearly independent.

b. Now suppose that $\operatorname{char} S = p \neq 0$. Since $d(x^p) = pd(x^{p-1}) = 0$ for any $x \in T$, we see that any S-linear derivation of T is automatically $S * T^p$-linear. Thus $\Omega_{T/S} = \Omega_{T/(S*T^p)}$, and it suffices to prove part b under the additional assumption that $T^p \subset S$. As remarked in Appendix 1, a p-basis for T over S is then just a minimal set of generators for T as an S-algebra.

In these circumstances if $\{x_\lambda\}_{\lambda\in\Lambda}$ is a p-basis then $T = S(\{x_\lambda\}_{\lambda\in\Lambda})$, and it follows that $\Omega_{T/S}$ is generated by the $\{dx_\lambda\}_{\lambda\in\Lambda}$. Suppose it were spanned by a proper subset, say all but dx_1. Let $T' := S(\{x_\lambda\}_{\lambda\in\Lambda,\lambda\neq1})$ as before. From the cotangent sequence it follows that $\Omega_{T/T'} = 0$. But if $x_1^p = y \in S$ then $T = T'[x]/(x^p - y)$, where x is an indeterminate, so by the conormal sequence $\Omega_{T/T'} = T\,dx/T \cdot d(x^p - y) = T\,dx$ is free on one generator, a contradiction.

Finally, supposing that $T^p \subset S$ and that $\{dx_\lambda\}_{\lambda\in\Lambda}$ is a basis of $\Omega_{T/S}$, we prove that $\{x_\lambda\}_{\lambda\in\Lambda}$ is a p-basis. We first show that $T = T' := S(\{x_\lambda\}_{\lambda\in\Lambda})$. From the cotangent sequence and the hypothesis that the $\{dx_\lambda\}_{\lambda\in\Lambda}$ generate $\Omega_{T/S}$, we see that $\Omega_{T/T'} = 0$. But if $\{y_\gamma\}_{\gamma\in\Gamma}$ were a p-basis of T over T', then the argument above shows that the $\{dy_\gamma\}_{\gamma\in\Gamma}$ are linearly independent elements of $\Omega_{T/T'}$, a contradiction.

To finish the argument we must show that $\{x_\lambda\}_{\lambda\in\Lambda}$ is a minimal set of generators for T. Otherwise, some x_λ, say x_1, would be in the subfield of T generated over S by the others, and so dx_1 would be in the subspace of $\Omega_{T/S}$ generated by the other dx_λ, a contradiction.

The last statement of the theorem follows because if T is finitely generated, then a p-basis is a separating transcendence basis by Corollary A1.5b. □

A number of related statements follow at once.

Corollary 16.16. *Suppose that S is a field. If T is a localization of a finitely generated S-algebra, then $\Omega_{T/S} = 0$ iff T is a finite direct product of fields, each finite and separable over S.*

For a Noetherian version valid in characteristic 0 see Exercise 16.11.

Proof. Suppose first that T is a finite direct product of fields finite and separable over S. By Proposition 16.10 it suffices to show that $\Omega_{T/S} = 0$ when T is itself a finite separable field extension of S, and this follows from Lemma 16.15 with $R = S$.

Now suppose that $\Omega_{T/S} = 0$. Since T is a localization of a finitely generated S-algebra it is Noetherian, and thus by Corollary 9.1 it will be enough to prove that each localization of T is a field, finite and separable over S. By Corollary 16.9, we may assume from the outset that T is itself a local ring.

Let $\mathfrak{m} \subset T$ be the maximal ideal. The conormal sequence

$$\mathfrak{m}/\mathfrak{m}^2 \to T/\mathfrak{m} \otimes_T \Omega_{T/S} \to \Omega_{(T/\mathfrak{m})/S} \to 0$$

shows that $\Omega_{(T/\mathfrak{m})/S} = 0$. By Theorem 16.14 and Corollary A1.5, $T/\mathfrak{m}$ is a finite separable extension of S. By Corollary 16.13 the map $\mathfrak{m}/\mathfrak{m}^2 \to T/\mathfrak{m} \otimes_T \Omega_{T/S}$ is an inclusion, so $\mathfrak{m}/\mathfrak{m}^2 = 0$. By Nakayama's lemma, $\mathfrak{m} = 0$ as well, and we are done.

Corollary 16.17. *Suppose $K \subset L$ are fields, with L finitely generated over K. Let r be the transcendence degree of L over K.*

a. $\dim_L \Omega_{L/K} \geq r$*, with equality iff L/K is separable.*

b. If L is separable over K, then any set of generators contains a separating transcendence base.

The example given in Exercise 16.10 shows that the hypothesis of finite generation is necessary in characteristic p (though it is not in characteristic 0).

Proof.

a. In characteristic 0 the result is immediate from Theorem 16.14. In characteristic p it follows from Theorem 16.14 because every p-basis contains a transcendence basis (and is equal to it in the finitely generated separable case) by Corollary A1.5.

b. Assume that L is separable over K. If $x_1, \ldots, x_n$ is a set of generators for L over K, then the dx_i generate $\Omega_{L/K}$. By Theorem 16.14 and Corollary A1.5 the subset $dx_1, \ldots, dx_r$ forms a basis iff $x_1, \ldots, x_r$ is a separating transcendence basis. But any set of generators for a vector space contains a basis. □

The next corollary is an improvement on the Noether normalization theorem.

Corollary 16.18 (Noether normalization with a separating transcendence basis). *If R is an affine domain of dimension d over a field k, and the quotient field L of R is separable over k, then there are elements $x_1, \ldots, x_d \in R$ such that R is integral over the polynomial ring $k[x_1, \ldots, x_d]$ (which is a subring) and also $x_1, \ldots, x_d$ form a separating transcendence base of L over k. If k is infinite, the x_i may be taken to be k-linear combinations of any given set of generators of R.*

Proof. If k is infinite, we may follow the proof of Theorem 13.3, according to which we should write $R = k[y_1, \ldots, y_r]/P$ for some prime P, and then modify the y_i to get x_i. Our d is the d_1 of the proof of Theorem 13.3. The elements $x_1, \ldots, x_d$ that are eventually produced are obtained by making a transformation of the form

$$x_i = y_i - \sum_{j>d} a_{ij} y_j, \qquad \text{with } a_{ij} \in k.$$

Now the condition for the x_i to be a separating transcendence base is that the elements dx_i generate $\Omega_{L/k}$, which, under our hypothesis, is a vector space of dimension d over L. Since the original dY_i must generate, it follows at once that the $x_1, \ldots, x_d$ will generate too if the coefficients a_{ij} are sufficiently general.

Without assuming that k is infinite, but allowing nonlinear changes of variable, we have seen in Lemma 13.2 that it suffices to make a transformation of the form

$$x_i = y_i - \sum_{j>d} y_j^{q^{e_j}},$$

for any sufficiently large integer q. If we take $q = p^N$ to be a power of the characteristic, then $dx_i = dy_i$, so it suffices to take the y_i to be a separating transcendence base to start with.

16.6 Jacobian Criterion for Regularity

Recall from Chapter 10 that a local ring $(R, \mathfrak{m})$ is **regular** iff $\mathfrak{m}$ can be generated by $\dim R$ elements. We asserted at that time that this notion has something to do with the geometric distinction between smooth and

singular points. In this section we shall prove an algebraic result that makes that relation clear and gives in many cases the most practical method for proving that an interesting ring is regular. It may be regarded as another version of the inverse function theorem.

Theorem 16.19 (Jacobian Criterion). *Let $S = k[x_1, \ldots, x_r]$ be a polynomial ring over a field k, let $I = (f_1, \ldots, f_s)$ be an ideal, and set $R = S/I$. Let P be a prime ideal of S containing I and write $\kappa(P) = K(R/P)$ for the residue class field at P. Let c be the codimension of I_P in S_P.*

a. The Jacobian matrix

$$\mathfrak{J} := (\partial f_i/\partial x_j),$$

taken modulo P, has rank $\leq c$.

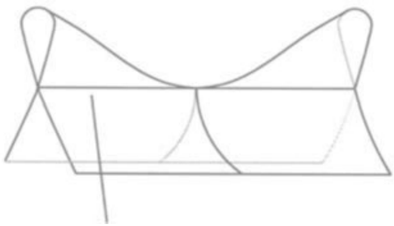
b. If $\operatorname{char} k = p > 0$, *assume that $\kappa(P)$ is separable over k. R_P is a regular local ring iff the matrix $\mathfrak{J}$, taken modulo P, has rank $= c$.*

We postpone the proof to describe some applications. The Jacobian criterion is often applied in the following special case. We say that an ideal has **pure codimension c** if all its minimal primes have codimension c.

Corollary 16.20. *Let $R = k[x_1, \ldots, x_r]/I$ be an affine ring over a perfect field k and suppose that I has pure codimension c. Suppose that $I = (f_1, \ldots, f_s)$. If J is the ideal of R generated by the $c \times c$ minors of the Jacobian matrix $(\partial f_i/\partial x_j)$, then J defines the singular locus of R in the sense that a prime P of R contains J iff R_P is not a regular local ring.*

The ideal $J \subset R$ (or its preimage in $k[x_1, \ldots, x_r]$, generated by the minors of the Jacobian matrix and the ideal I, for which we shall also write J) is called the **Jacobian ideal** of R. It follows from Corollary 16.20 that the radical of J depends only on k and R, and not on the chosen presentation $R = k[x_1, \ldots, x_r]/I$ of R as a homomorphic image of a polynomial ring over k. But in fact the ideal J itself depends only on k and R : J is the dth fitting ideal of $\Omega_{R/k}$, where $d = r - c$ is the dimension of R. See Chapter 20 for the definition and basic properties of fitting ideals.

As an example of how this works, consider the surface X defined by the equation $y^2 - x^2(t^2 - x) = 0$ in $\mathbf{A}^3$, over a field k. The function t on X has level sets that are nodal curves when $t \neq 0$, degenerating to a cuspidal curve when $t = 0$, as in Figure 16.3. The Jacobian ideal is $J = (y^2 - x^2(t^2 - x), 3x^2 - 2t^2x, 2y, 2tx^2)$, which has radical (x, y). A prime P of R contains J iff P contains (x, y) iff $P = (x, y)$ or $P = (x, y, f(t))$, where f is an irreducible polynomial over k. If char $k \neq 2$ then the associated primes of J are (x, y) and (x, y, t) and this is one way to see algebraically that something different is happening at the prime (x, y, t); from the picture one sees immediately that the corresponding point $x = y = t = 0$ is "more" singular than the others. Quite a lot of refined information is available from



$k[x,y,t]/(y^2 - x^2(t^2 - x))$.
$I = (y^2 - x^2(t^2 - x))$; codim $= 1$.
$\mathrm{J} = (3x^2 - 2t^2x, 2y, -2tx^2)$.
The ideal of 1×1 minors, plus I, is
$J = (y^2 - x^2(t^2 - x), 3x^2 - 2t^2x, 2y, 2tx^2)$,
which has radical (x, y)

The singular locus is
$x = y = 0$, a line
parametrized by t.

FIGURE 16.3.

the Jacobian ideal and related constructions; see, for example, Morin [1975] and Lê and Teissier [1981] for more information.

Combined with Serre's criterion for normality, the Jacobian criterion is very useful for proving that certain rings are normal, or even in checking that certain ideals are prime; see Theorem 18.15 and the discussion there.

Useful as the Jacobian criterion is, it has two technical limitations: The ground field k must be perfect, and the criterion deals only with localizations of affine algebras. One certainly wants to know, and it is true, that there is an ideal defining the singular locus of a finitely generated algebra over an nonperfect field, or over the integers. For general Noetherian rings, however, there is no such ideal: The set of primes P such that R_p is not regular need not in general be a closed set in the Zariski topology! Grothendieck set up a very elegant framework, the theory of "smoothness" of a ring R with respect to an ideal I, through which the limitations of the criterion may, to a certain extent, be overcome; see, for example, Matsumura [1986, Chapter 10] for an exposition.

Proof of Theorem 16.19.

a. Pulling back P to $k[x_1, \ldots, x_r]$ we may regard P as a prime in the polynomial ring. Let Q be a prime of $k[x_1, \ldots, x_r]$ having codimension c and satisfying $I \subset Q \subset P$. It suffices to show that the rank of the Jacobian matrix modulo Q is at most c, and of course it suffices to do this after extending the set of f_i to generate Q; that is, we may assume $I = Q = P$. The conormal sequence

$$I/I^2 \to R \otimes \Omega_{k[x_1,\ldots,x_r]/k} \to \Omega_{R/k} \to 0,$$

for $k \to k[x_1, \ldots, x_r] \to R$, localized at Q, shows that the cokernel of the Jacobian matrix, regarded as a matrix over the field $\kappa(Q) = R_Q$, is the module $(\Omega_{R/k})_Q$, which by Proposition 16.9 is $\Omega_{R_Q/k}$. By Corollary 16.17, $\dim_{R_Q} \Omega_{R_Q/k} \geq r - c$. Since the Jacobian matrix has r rows, its rank, taken modulo Q, is $\leq c$.

b. Now suppose that $\kappa(P)$ is separable over k. The conormal sequence for $k \to R_P \to \kappa(P)$ is

$$P_P/P_P^2 \xrightarrow{d} \kappa(P) \otimes \Omega_{R_P/k} \to \Omega_{\kappa(P)/k} \to 0.$$

By Corollary 16.13 the map d is an injection, so

$$\dim_{\kappa(P)} P_P/P_P^2 + \dim_{\kappa(P)} \Omega_{\kappa(P)/k} = \dim_{\kappa(P)} \kappa(P) \otimes \Omega_{R_P/k}.$$

Now $\dim_{\kappa(P)} P_P/P_P^2 \geq \dim R_P$, with equality iff R_P is regular. By Corollary 16.17a, $\dim_{\kappa(P)} \Omega_{\kappa(P)/k} =$ tr. deg. $\kappa(P)/k$, since $\kappa(P)$ is separably generated over k. By Theorem A of Chapter 8 (proven in Chapter 13), we can rewrite tr. deg. $\kappa(P)/k$ as $\dim R/P$, and $\dim R/P + \dim R_P = r - c$. Putting these facts together we get

$$\dim_{\kappa(P)} \kappa(P) \otimes \Omega_{P/k} \geq r - c$$

with equality iff R_P is regular.

It remains to connect $\dim_{\kappa(P)} \kappa(P) \otimes \Omega_{R_P/k}$ to the rank of the Jacobian matrix. If we map R^s onto I by sending the ith generator to f_i then, as in the section on computation of differentials, we may regard the Jacobian matrix $\mathcal{J} = (\partial f_j/\partial x_i)$ as the composite

$$R^s \to I/I^2 \to R \otimes \Omega_{k[x_1,\ldots,x_r]/k} = R^r.$$

The conormal sequence

$$I/I^2 \to R \otimes \Omega_{k[x_1,\ldots,x_r]/k} \to \Omega_{R/k} \to 0,$$

for $k \to k[x_1, \ldots, x_r] \twoheadrightarrow R$ shows that $\Omega_{R/k}$ is the cokernel of the matrix $\mathcal{J}$, and thus $\kappa(P) \otimes \Omega_{R_P/k}$ is the cokernel of the matrix $\mathcal{J}$ taken modulo P. It follows that $\dim_{\kappa(P)} \kappa(P) \otimes \Omega_{R_P/k} = r - c$ iff the rank of the Jacobian matrix, taken modulo P, is c. □

16.7 Smoothness and Generic Smoothness

For the following applications we need the connection between locally free modules and the ranks of matrices. This simple but important idea will be studied in detail in Chapter 20 under the rubric of fitting ideals. For the reader's convenience we explain it briefly here. Suppose that $\mathcal{J} : R^t \to R^r$ is a map of free modules over a ring R whose rank is $\leq c$, as for the Jacobian matrix of an ideal of codimension c (Theorem 16.19), and let $M = \operatorname{coker} \mathcal{J}$. Let P be a prime ideal of R. We claim that M_P is free of rank $r - c$ iff the matrix $\mathcal{J}$, taken modulo P, has rank exactly c (that is, some $c \times c$ minor of $\mathcal{J}$ is outside P). This is a special case of Proposition 20.8, but here in a nutshell is the argument:

We may as well assume that R is local with maximal ideal P to start with. Suppose first that M is free. Tensoring the sequence with R/P, we may assume that R is a field; the rank of the vector space M/PM is $r-c$ because M was supposed free. It follows that the rank of the map $\mathfrak{J} \otimes R/P$ is c, and this is the desired conclusion.

Conversely, suppose that the rank of $\mathfrak{J}$, taken modulo P, is c. One of the $(r-c) \times (r-c)$ submatrices of $\mathfrak{J}$ must have determinant outside of P, and thus be invertible. Multiplying $\mathfrak{J}$ by a suitable invertible matrix (an operation that does not change the freeness of the cokernel of $\mathfrak{J}$) we may suppose that $\mathfrak{J}$ has the form

$$\mathfrak{J} = \left(\begin{array}{ccccc|c} 1 & 0 & 0 & \dots & 0 & \\ 0 & 1 & 0 & \dots & 0 & \mathfrak{J}_{12} \\ & \cdot & \cdot\cdot & \cdot\cdot & & \\ 0 & 0 & \dots & 0 & 1 & \\ \hline & & \mathfrak{J}_{21} & & & \mathfrak{J}_{22} \end{array}\right)$$

where the upper left-hand corner is an $(r-c) \times (r-c)$ identity matrix. Performing row and column operations (which again amounts to multiplying $\mathfrak{J}$ on the left and right by invertible matrices, not changing the freeness of coker $\mathfrak{J}$), we may reduce to the case where $\mathfrak{J}_{12}$ and $\mathfrak{J}_{21}$ are the zero matrices. From the fact that rank $\mathfrak{J} = r - c$, we see then that $\mathfrak{J}_{22}$ must be a zero matrix as well. It is now obvious that the cokernel of $\mathfrak{J}$ is free of rank c.

Corollary 16.21. *Let $S = k[x_1, \dots, x_r]$ be a polynomial ring over a field k, let $I \subset S$ be an ideal, and set $R = S/I$. Let P be a prime ideal of S containing I whose residue class field $\kappa(P) = K(R/P)$ is separable over k, and let $c = \operatorname{codim} I_P \subset S_P$. R_P is a regular local ring iff the module $\Omega_{R/k}$ is locally free at P of rank $r - c$.*

Proof. Localizing the conormal sequence, we get a free presentation

$$R_P^t \xrightarrow{\mathfrak{J}} R_P \otimes \Omega_{S/k} \to (\Omega_{R/k})_P \to 0.$$

Thus by the argument given above, $(\Omega_{R/k})_P$ is free of rank $r-c$ iff $\mathfrak{J}$ has rank c locally at P.

Note that in the common situation where I is a prime ideal, the number $r - c$ may be computed directly from R_P as the transcendence degree of $K(R/P)$ over k.

It is not true that if $\Omega_{R/k}$ is locally free at P then R_P is regular; in characteristic p the rank could be $> r-c$. See Exercise 16.7 for an example. But all is well over a perfect field.

Corollary 16.22. *Let k be a perfect field, and let R be an affine ring over k. Suppose that R is equidimensional, of dimension d. The module $\Omega_{R/k}$ is*

locally free over R *of rank* d *iff* R_P *is a regular local ring for each prime* P *of* R. *If* R *is reduced or* $\operatorname{char} k = 0$, *then the condition that the rank of* $\Omega_{R/k}$ *is* d *is automatically satisfied.*

In fact, the proof below works whenever the characteristic of k is larger than the degree of nilpotence of P_P for any minimal prime P.

Proof. Since every finitely generated extension of a perfect field is separably generated by Theorem A1.3, everything but the last statement follows from Corollaries 16.21 and 16.17. Suppose that $\Omega_{R/k}$ is locally free over R; we must show that if R is reduced or $\operatorname{char} k = 0$, then $\Omega_{R/k}$ has rank $d = \dim R$. Let P be a minimal prime of R, so that $\dim R/P = \dim R = d$. The quotient field $\kappa(P) = K(R/P)$ has transcendence degree d over k by Theorem A of Chapter 13. The conormal sequence

$$P_P/P_P^2 \xrightarrow{d} \kappa(P) \otimes \Omega_{R/k} \to \Omega_{\kappa(P)/k} \to 0$$

shows that the rank of $\Omega_{R/k}$ differs from the transcendence degree of $\kappa(P)$, which is d, by the dimension of the image of the map $d : P_P/P_P^2 \to \kappa(P) \otimes \Omega_{R/k}$. Thus it will suffice to show that $df = 0$ in $\kappa(P) \otimes \Omega_{R/k}$ for any $f \in P_P$. If R is reduced, then $P_P = 0$, so we may assume that $\operatorname{char} k = 0$.

If $f \in P_P$ then f is nilpotent in R_P; say $f^n = 0$ but $f^{n-1} \neq 0$. Applying d, we get $nf^{n-1}df = 0$ in $\Omega_{R_P/k}$, and since $\operatorname{char} k = 0$ we get $f^{n-1}df = 0$. Since $\Omega_{R_P/k} = (\Omega_{R/k})_P$ by Proposition 16.9, $\Omega_{R_P/k}$ is a free R_P-module. If $df \notin P\Omega_{R_P/k}$, then df would generate a free summand of $\Omega_{R_P/k}$. Since df is annihilated by the nonzero element f^{n-1}, we see that this is impossible, and $df \in P\Omega_{R_P/k}$. Thus $df = 0$ in $\kappa(P) \otimes \Omega_{R_P/k} = \kappa(P) \otimes (\Omega_{R/k})$, as required.

We shall say that an equidimensional affine ring R over a perfect field k is **smooth** over k if $\Omega_{R/k}$ is locally free of rank $= \dim R$.

A crucial result in the theory of manifolds is Sard's theorem, which says that if $\varphi : X \to Y$ is a $\mathcal{C}^\infty$ map of smooth manifolds, and $X' \subset X$ is the critical locus of φ (the locus where the derivative of φ drops rank), then $\varphi(X')$ has measure 0 in Y. The following is the algebraic version, which shows that if X and Y are varieties over a field of characteristic 0, then $\varphi(X')$ is contained in a proper subvariety of Y.

Corollary 16.23 (Generic smoothness). *Let* $R \subset S$ *be affine domains over a perfect field* k. *Suppose that* $K(R) \otimes S$ *is smooth over* $K(R)$, *or that* $\operatorname{char} k = 0$ *and* S_Q *is a regular local ring for every maximal ideal* Q. *There is an element* $0 \neq f \in R$ *such that the fiber* $\kappa(P) \otimes_R S$ *is smooth over* $\kappa(P)$ *for every prime* P *of* R *not containing* f.

Proof. Let $T = K(R) \otimes_R S$. Since the formation of differentials localizes (Proposition 16.9), we have $\Omega_{T/K(R)} = K(R) \otimes_S \Omega_{S/R}$.

Suppose that $\operatorname{char} k = 0$ and S_Q is a regular local ring for every maximal ideal Q. Now T is an affine ring over $K(R)$ whose localizations are all localizations of S, and are thus regular. It follows that $\Omega_{T/K(R)}$ is locally

free over T of rank equal to the dimension of T, which is the transcendence degree of $K(S)$ over $K(R)$, or equivalently $\dim S - \dim R$. Thus in both cases of the theorem, $K(R) \otimes S$ is smooth over $K(R)$.

Since $\Omega_{T/K(R)}$ is free over T of rank $\dim T - \dim K(R)$, we may find an element $0 \neq f_1 \in R$ such that setting $R' = R[f_1^{-1}]$ and $S' = S[f_1^{-1}]$, the module $\Omega_{S'/R'}$ is free of rank $\dim S' - \dim R'$ over S'.

By base change, Proposition 16.4, we see that $\Omega_{\kappa(P)\otimes_R S/\kappa(P)} = \kappa(P) \otimes_R \Omega_{S'/R'} = (\kappa(P) \otimes_R S') \otimes_{S'} \Omega_{S'/R'}$ for every prime P of R not containing f_1. Thus $\Omega_{\kappa(P)\otimes_R S/\kappa(P)}$ is free over $\kappa(P)\otimes_R S = \kappa(P)\otimes_R S'$ of rank $\dim S - \dim R$ for every such prime.

By the semicontinuity of fiber dimension, Corollary 14.6, there is an element $f_2 \in R$ such that $\kappa(P) \otimes_R S$ has dimension equal to $\dim S - \dim R$ for every prime P of R not containing f_2. If we take $f = f_1 f_2$ we see that $\kappa(P) \otimes_R S$ is smooth over $\kappa(P)$ for every prime P of R not containing $f_1 f_2$, as required.

For a simple application, consider a polynomial $g(x_1, \ldots, x_r) \in k[x_1, \ldots, x_r]$. If k has characteristic 0, then Corollary 16.23 implies that for all but finitely many values of $t \in k$ the ring $k[x_1, \ldots, x_r]/(g - t)$ is smooth over k. (If we take $R = k[y]$ mapping to $S = k[x_1, \ldots, x_r]$ by sending y to g, then this is true for any t that is not a root of the polynomial f produced in the corollary.) Note that this fails in characteristic p: If k is perfect, for example, then $t^{1/p}$ is in k for every $t \in k$, so taking $g(x) = x_1^p$ we see that $g - t = (x_1 - t^{1/p})^p$ for every $t \in k$, and $k[x_1, \ldots, x_r]/(g - t)$ is not even regular.

16.8 Appendix: Another Construction of Kähler Differentials

There is another construction of the module of derivations that is too important to ignore, though we shall not use it in this book:

Theorem 16.24. *Let I be the kernel of the multiplication map $\mu : S \otimes_R S \to S$. If $e : S \to I/I^2$ is the map defined by $b \mapsto 1 \otimes b - b \otimes 1$, then there is an isomorphism $\varphi : \Omega_{S/R} \to I/I^2$ of S-modules such that $\varphi d = e$; that is, the pair $(d, \Omega_{S/R})$ is in a natural sense isomorphic to $(e, I/I^2)$.*

Theorem 16.24 essentially writes $\Omega_{S/R}$ as the conormal module for the map of rings $S \otimes_R S \to S$. If R and S are affine rings and the map $R \to S$ corresponds to a map of affine varieties $Y \to X$, then the epimorphism $S \otimes_R S \to S$ corresponds to the diagonal embedding of Y in $Y \times_X Y$.

To understand why this works, consider a smooth manifold M. The normal bundle of the diagonal embedding of M in $M \times M$ is the cokernel of the diagonal map

$$T_M \to T_{M\times M|M} = T_M \oplus T_M$$

and is thus T_M. Thus the cotangent bundle of M is the conormal bundle of this diagonal embedding. Theorem 16.24 is the general algebraic form of this assertion.

Theorem 16.24 is significant for two reasons. First, it is a special case of a more general construction, extending the idea of the module of differentials: The ring $S \otimes_R S/I^{n+1}$ may be viewed as an S-module by the map $S \to S \otimes_R S/I^{n+1}$ sending b to $b \otimes 1$. If $R = k$ is a field and S is a localization of the affine ring of a nonsingular variety X over k, then this module is the module of sections of the jet bundle of order n on X: That is, it parametrizes Taylor series expansions to order n of polynomial functions on X.

Second, if you know about the global treatment of varieties and schemes by means of sheaf theory, you will see that this construction, unlike the one given at the beginning of this chapter, "globalizes without patching": It makes sense directly for sheaves of functions, as well as for rings.

Proof. We first show that e is a derivation. Indeed, e is the difference of the two maps $b \mapsto 1 \otimes b$ and $b \mapsto b \otimes 1$, which are algebra maps splitting the sequence

$$I/I^2 \to (S \otimes_R S)/I^2 \to S \to 0.$$

Quite generally, if $T \to S$ is an algebra map whose kernel J has square 0, then J is naturally an S-module. By Proposition 16.11 any two splittings $e_1, e_2 : S \to T$ differ by a derivation.

Because of the universal property of d and $\Omega_{S/R}$, there is a unique map $\varphi : \Omega_{S/R} \to I/I^2$ satisfying $e = \varphi d$; that is, with $\varphi(db) = 1 \otimes b - b \otimes 1$. It remains to prove that φ is an isomorphism, which we shall do by identifying its inverse. Let $T := S \ltimes \Omega_{S/R}$ be the "trivial extension of S by $\Omega_{S/R}$"; that is, the ring which, as an abelian group, is the direct sum of S and $\Omega_{S/R}$, and whose multiplication is defined, for $b, b' \in S$, and u, u' in $\Omega_{S/R}$ by

$$(b, u)(b', u') = (bb', bu' + b'u).$$

We claim that there is a ring homomorphism

$$\psi : S \otimes_R S \to T; \qquad \psi : a \otimes b \mapsto (ab, adb) \quad \text{for } a, b \in S.$$

Since $\psi(1 \otimes b - b \otimes 1) = (0, db)$, we see that the restriction of ψ to I induces the desired inverse on I/I^2.

A map from the tensor product $S \otimes_R S$ to an R-algebra T may be specified by giving two maps of R-algebras $\psi_i : S \to T$, so it is enough to show that

$$\begin{aligned} \psi_1 : S \to T; \qquad & \psi_1 : b \mapsto (b, db) \quad \text{for } b \in S \\ \psi_2 : S \to T; \qquad & \psi_2 : a \mapsto (a, 0) \quad \text{for } a \in S \end{aligned}$$

are maps of R-algebras. This is immediate for ψ_2, and follows because d is an R-linear derivation in the case of ψ_1, so we are done.

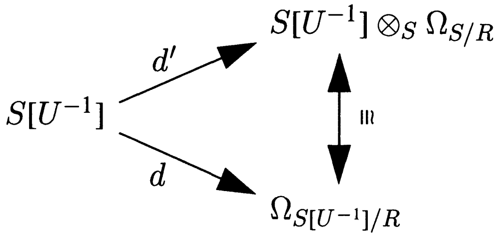
16.9 Exercises

Exercise 16.1:* Show that if $b \in S$ is an idempotent (that is, $b^2 = b$), and if $d : S \to M$ is any derivation, then $db = 0$.

Exercise 16.2:* Let M be an R-module, and let $S = R \ltimes M$, the "trivial extension of S by M"; that is, as an R-module $S = R \oplus M$, and the multiplication is the obvious one, with $M^2 = 0$. Compute $\Omega_{S/R}$ and the universal derivation.

Exercise 16.3: Let $S = R[x,y]/(xy)$. Compute $\Omega_{S/R}$.

Exercise 16.4: Give a direct proof, without using exact sequence and colimits, of Proposition 16.9 (localization of differentials), as follows. If R is an S-algebra and U is a multiplicatively closed subset of S, show that there is an R-linear derivation $d' : S[U^{-1}] \to S[U^{-1}] \otimes_S \Omega_{S/R}$ sending $1/s$ to $s^{-2}ds$, and a commutative diagram



To define the upward map, use the fact that the obvious composite map $S \to S[U^{-1}] \to \Omega_{S[U^{-1}]/R}$ is a derivation. For the downward map, one must check that $d(b/s) = (1/s^2)(sdb - bds)$ is legitimate. (After some computation this boils down to showing that if $s \in U$ kills b (so that $b/t = 0$ for all $t \in S$), then s^2 kills db (so that $d(b/t) = 0$).)

Exercise 16.5: Let $(S, \mathfrak{m})$ be a regular local ring that is the localization at a maximal ideal of a finitely generated algebra over a field k, and let $x_1, \ldots, x_d$ be a system of parameters. Show that if $S/\mathfrak{m} = k$, or more generally if $S/\mathfrak{m}$ is a separable extension of k, then $\Omega_{S/k}$ is a free S-module of rank d, generated by the dx_i. What about the case when $S/\mathfrak{m}$ is not separable over k?

Exercise 16.6: If $K \subset L' \subset L$ are fields finitely generated over K, then the cotangent sequence gives a natural map

$$\varphi : L \otimes \Omega_{L'/K} \to \Omega_{L/K},$$

and Lemma 16.15 shows that this is an isomorphism if L is separable and algebraic over L'.

a. Show that if $L = L'(x_1, \ldots, x_r)$ is the field of rational functions in $r > 0$ indeterminates, then φ is injective but not surjective.

b. In general φ need not be injective. Construct an example as follows: Suppose that L is purely inseparable and algebraic over L', generated by one element α satisfying a minimal polynomial $f(x) = x^p - a$, with $a \in L'$. By Corollary 16.6, the conormal sequence for $K \to L'[x] \to L$ has the form

$$Ldf \to L \otimes \Omega_{L'/K} \oplus Ldx \to \Omega_{L/K} \to 0.$$

Show that φ is injective iff $da = 0 \in \Omega_{L'/K}$. Construct an explicit example where this condition is not satisfied.

Exercise 16.7: Let k be a field of characteristic p, and let $R = k[x]/x^p$. Show that $\Omega_{R/k}$ is free although R is not regular. Note that the rank of $\Omega_{R/k}$ is not equal to the dimension of R (compare with Corollary 16.21).

Exercise 16.8 (First-order deformations): We say that deformations $\tilde{R}$ and $\tilde{R}'$ with base ring A are **isomorphic** if there is an isomorphism $\alpha : \tilde{R} \to \tilde{R}'$ of A-algebras such that the induced map $\alpha \otimes_A k : \tilde{R} \otimes_A k \to \tilde{R}' \otimes_A k$ corresponds under the given isomorphisms $R \cong \tilde{R} \otimes_A k$ and $R \cong \tilde{R}' \otimes_A k$ to the identity map of R. If A is a k-algebra in such a way that the composite map $k \to A \to k$ is the identity, then $\tilde{R} \cong A \otimes_k R$ is a deformation, called the **trivial** deformation. If $A = k[x]/(x^2)$, then a deformation with base ring A is called a **first-order infinitesimal deformation** of R over k.

Suppose that k is a field, and that R is an affine ring; say $R = S/I$, where $S := k[x_1, \ldots, x_r]$ is a polynomial ring. In Exercise 6.12 we saw that the set of first-order infinitesimal embedded deformations of R is in one-to-one correspondence with $N := \operatorname{Hom}(I/I^2, R)$. In this exercise, we shall compute the set of all first-order deformations, which turns out to be a quotient of N. The reader will need to use the results of Exercise 6.12 here.

As with the case of embedded deformations, the set of first-order infinitesimal deformations of R over k is in a natural sense the Zariski tangent space to the space of all deformations. We refer the reader to Eisenbud and Harris [1992] for more information.

a. Suppose that $\tilde{R}$ and $\tilde{R}'$ are first-order infinitesimal deformations of the k-algebra R, and that there is a map of $k[\varepsilon]$-algebras $\alpha : \tilde{R} \to \tilde{R}'$ such that the induced map

$$R \cong \tilde{R} \otimes_{k[\varepsilon]} k \xrightarrow{\alpha \otimes 1} \tilde{R}' \otimes_{k[\varepsilon]} k \cong R$$

is the identity map from R to R. Show that α is an isomorphism between $\tilde{R}$ and $\tilde{R}'$, so that the deformations are isomorphic. For example let $S = k[x]$, $R = k[x]/(x^n)$. Show that the deformation $S[\varepsilon]/(x - a\varepsilon)^n = S/(x^n - na\varepsilon x^{n-1})$ of R is trivial as a (though it is not trivial as an embedded deformation.)

b. Show that every first-order infinitesimal deformation of R comes from an embedded first-order infinitesimal deformation, as follows: Suppose that $S = k[x_1, \ldots, x_r], R = S/I$, and $\tilde{k}[\varepsilon] \to \tilde{R}$ is a deformation of R. Let y_i be the image of x_i in R, and let $\tilde{y}_i$ be any element of $\tilde{R}$ whose image in R is y_i. Show that $\tilde{R}$ is generated over $k[\varepsilon]$ by the $\tilde{y}_i$, so that there is a surjection $S[\varepsilon] \twoheadrightarrow \tilde{R}$ "embedding" the given deformation.

c. Again, let $S = k[x_1, \ldots, x_r]$, and $R = S/I$. Let $\tilde{R} = S[\varepsilon]/\tilde{I}$ and $\tilde{R}' = S[\varepsilon]/\tilde{I}'$ be two first-order infinitesimal embedded deformations that are isomorphic in the sense above, with isomorphism α. Show that α can be lifted to a map $\tilde{\alpha} : S[\varepsilon] \to S[\varepsilon]$ of $k[\varepsilon]$-algebras of the form

$$\tilde{\varphi} : x_i \mapsto x_i + \varepsilon a_i \quad \text{with } a_i \in S.$$

Show that for any $\tilde{g} = g(x) + \varepsilon g_1(x) \in S[\varepsilon], \tilde{\alpha}(\tilde{g}) = \tilde{g} + \varepsilon \sum a_i \partial g/\partial x_i$ (note that this is an "infinitesimal Taylor expansion" of g).

d. With notation as in part c, suppose that the two given deformations correspond in the sense of Exercise 6.12 to homomorphisms $\varphi, \varphi' : I/I^2 \to R$. Show that $\varphi - \varphi' =$ takes g to $\sum a_i \partial g/\partial x_i$. Conversely, show that if $\varphi, \varphi' : I/I^2 \to R$ are any two homomorphisms, differing by a homomorphism of the form $g \mapsto \sum a_i \partial g/\partial x_i$, then φ and φ' define embedded first-order deformations that are isomorphic as (nonembedded) first-order deformations.

e. Consider the conormal sequence of $k \to S \twoheadrightarrow R$,

$$I/I^2 \to R \otimes \Omega_{S/k} \to \Omega_{R/k} \to 0.$$

Dualizing into R we get a left-exact sequence

$$0 \to \operatorname{Hom}(\Omega_{R/k}, R) \to \operatorname{Hom}(R \otimes \Omega_{S/k}, R) \to \operatorname{Hom}(I/I^2, R).$$

Define $T^1_{R/k}$ to be the cokernel of the map $\operatorname{Hom}(R \otimes \Omega_{S/k}, R) \to \operatorname{Hom}(I/I^2, R)$ in this sequence. Show that $T^1_{R/k}$ is the set of isomorphism classes of first-order infinitesimal deformations of R over k.

f. (Compare with Exercise 6.12e) Let $S = k[x], R = k[x]/(x^n)$ (the n-fold point on a line"). Show that each first-order (nonembedded) deformation may be written in the form

$$S[\varepsilon]/(x^n + a_2 \varepsilon x^{n-2} + \cdots + a_n \varepsilon)$$

for a unique $a_2, \ldots, a_n \in k$. Note that there is no term in x^{n-1}. Geometrically this corresponds to a family of n points on a line approaching 0, modulo translations.

g. (Compare with Exercise 6.12f) Let $S = k[x, y], R = k[x]/(xy)$ (the "ordinary double point"). Show that each first-order (nonembedded)

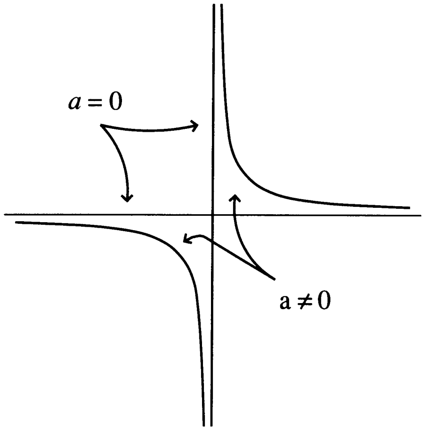


FIGURE 16.4.

deformation may be written in the form $S[\varepsilon]/(xy + a\varepsilon)$ for a unique $a \in k$. Note that the space of these deformations is just 1-dimensional over k. If $k = \mathbf{R}$, this deformation corresponds to the tangent vector to the deformation of two lines meeting at a point into a hyperbola, as shown in Figure 16.4.

Exercise 16.9: Use Theorem 16.8 to prove that if R is a field and S is a separable algebraic extension of R (possibly not finitely generated) then $\Omega_{S/R} = 0$.

Exercise 16.10: Let S be a field of characteristic $p > 0$, let x be an indeterminate, and set

$$T = S(x^{1/p^\infty}) = \varinjlim S(x^{1/p^n}).$$

Show directly that $\Omega_{T/S} = 0$, although T is not separably generated over S; this shows that the hypothesis of finite generation is necessary in Corollary 16.17 in characteristic p.

Exercise 16.11: Let S be a field of characteristic 0, and let T be a Noetherian S-algebra such that $\Omega_{T/S} = 0$. Show that T is a finite direct product of algebraic field extensions of S.

Exercise 16.12: Let k' be a field of characteristic p, and let $K = k'(t), S = k'(t)[x]/(x^2)$ (or if you prefer take $S = k'(t)[[x]]$). Let $k = k'(t^p + x) \subset S$. Show that the map d: $\mathfrak{m}/\mathfrak{m}^2 \to K \otimes \Omega_{S/k}$ is zero. (Compare with Theorem 7.8 and Exercise 7.17.)

Exercise 16.13:* Let $S = k[x_1, \ldots, x_r]$ be a polynomial ring over a field k, let $I = (f_1, \ldots, f_s)$ be an ideal, and set $R = S/I$. Suppose I has codimension c. The first part of the proof of Theorem 16.19 shows that $(c+1) \times (c+1)$ minors of the Jacobian matrix $(\partial f_i/\partial x_j)$ are contained in any prime of

codimension c that contains I (by the Nullstsellensatz this is equivalent to the first statement of Theorem 16.19). Show by example that the minors need not be contained in I itself, even if I is unmixed.

Exercise 16.14 (The complete case):* In the complete case the module of differentials as we have defined it is not so useful. For example, if R is the localization of an affine ring over a field k at a maximal ideal and $\hat{R}$ is its completion, then in general $\Omega_{\hat{R}/k} \neq \hat{R} \otimes_R \Omega_{R/k}$. For most applications (such as the Jacobian criterion) it is $\hat{R} \otimes_R \Omega_{R/k}$ that is interesting. It turns out that this is the completion of $\Omega_{\hat{R}/k}$, and that in general the completed module of differentials is the right thing. Here is the beginning of the treatment; see Grothendieck [1964, 20.4.8.2] and Scheja and Storch [1972] for continuations that also include the case of rings of germs of analytic functions (convergent power series).

a. Let $= \mathbf{Q}[[x_1, \ldots, x_r]]$ be the ring of formal power series in $r \geq 1$ variable over the rational numbers. Show that $\Omega_{R/\mathbf{Q}}$ is not a finitely generated—or even a countably generated—R-module.

b. Let (S, P) be a complete local ring with coefficient field k. Show that the completion $\hat{\Omega}_{S/k} := \varprojlim_n (\Omega_{S/k})/P^n\Omega_{S/k}$ may be identified with the inverse limit of the modules $\Omega_{(S/P^n)/k}$. Show that $\hat{\Omega}_{S/k} = (\Omega_{S/k})/\cap_{j=1}^{\infty} P^j\Omega_{S/k}$. Show that the natural derivations $S \to S/P^n \to \Omega_{(S/P^n)/k}$ give rise to a derivation $\hat{d} : S \to \hat{\Omega}_{S/k}$, which may also be identified as the composite of the universal derivation $d : S \to \Omega_{S/k}$ and the projection map $\Omega_{S/k} \to (\Omega_{S/k})/\cap_{j=1}^{\infty} P^j\Omega_{S/k}$.

c. If $R = k[[x_1, \ldots, x_r]]$ is the formal power series ring, show that $\hat{\Omega}_{R/k}$ is the free R-module generated by $dx_1, \ldots, dx_r$.

d. Let (S, P) be a complete local ring with coefficient field k, where P is generated by elements $y_1, \ldots, y_r$. Show that if we write S as $k[[x_1, \ldots, x_r]]/(f_1, \ldots, f_t)$, then

$$\hat{\Omega}_{S/k} = (\oplus_i R dy_i)/S df_1 + \cdots + S df_t$$

exactly as in the affine case.

Exercise 16.15 (The de Rham complex):

a. Let $\Omega^i_{S/R} = \wedge^i \Omega_{S/R}$, the ith exterior power as S-modules. Show that the universal derivation $d : S = \Omega^0_{S/R} \to \Omega^1_{S/R} = \Omega_{S/R}$ is in fact the first step in a complex of R-modules

$$\Omega^\bullet_{S/R} : 0 \to \Omega^0_{S/R} \xrightarrow{d} \Omega^1_{S/R} \xrightarrow{d^1} \cdots \to \Omega^i_{S/R} \xrightarrow{d^i} \Omega^{i+1}_{S/R} \to \cdots .$$

where the map $d^i : \Omega^i_{S/R} \to \Omega^{i+1}_{S/R}$ satisfies

$$d^i(bdb_1 \wedge db_2 \wedge \cdots \wedge db_i) = db \wedge db_1 \wedge db_2 \wedge \cdots \wedge db_i.$$

(You must check that the right side depends only on the left side, and, given this, that the formula defines a map of R-modules.) The complex $\Omega^\bullet_{S/R}$ is called the **de Rham complex** of S relative to R.

The de Rham complex has long been used in the theory of manifolds to compute topological cohomology; see, for example, Bott and Tu [1982]. Atiyah, Hodge, and Grothendieck observed that in the algebraic setting, if $R = \mathbf{C}$, and S is the affine ring of a smooth affine variety X, then the homology of the de Rham complex at $\Omega^i_{S/R}$ is $H_i^{\text{sing}}(X, \mathbf{C})$, the usual singular homology group; see Hartshorne [1975] (where the history is also surveyed). The following examples illustrate these ideas:

b. If $R = k$ is a field, and S is the affine ring of a variety X of dimension d, then we say that X is **smooth over R** if $\Omega_{S/R}$ is locally free of rank d as an S-module. In this case, show that $\Omega^i_{S/k} = 0$ for $i > d$. This corresponds to the fact that although X has dimension $2d$ as a real manifold, X has the homotopy type of a complex of dimension $\leq d$. (See Andreotti and Frankel [1959] for a proof.)

c. Let $S = R[x_1, \ldots, x_r]$ be the polynomial ring in r variables. Show that the de Rham complex of S/R is exact except at $\Omega^0_{S/R}$, where the homology is R. (Note that this is *not* a complex of S-modules.) This example corresponds to the fact that the only nonvanishing homology group of affine space is H_0.

d. Let $R = k$ be a field, and let $S = k[x, y]/(f(x, y))$, where $f(x, y) = y(x - a_1)(x - a_2)\cdots(x - a_d) - 1$ where $a_1, \ldots, a_d \in k$ are distinct. Show that $\Omega_{S/R}$ is free of rank 1 with generator dx. Show that the homology of the de Rham complex

$$0 \to S \to \Omega_{S/R} \to 0$$

is k in degree 0 and k^d in degree 1. (This corresponds to the fact that S is the coordinate ring of the smooth affine variety $\mathbf{A}^1 - \{d \text{ points}\}$, which, if $k = \mathbf{C}$, has the homotopy type of a bouquet of d circles.)

e. Let $R = k$ be a field of characteristic not 2 or 3, and let $S = k[x, y]/(x^3 + y^2 - 1)$. Check that $\Omega_{S/R}$ is a locally free module of rank 1, so the de Rham complex is again

$$0 \to S \to \Omega_{S/R} \to 0.$$

Show by finding k-bases for the two (infinite-dimensional) vector spaces S and $\Omega_{S/R}$ that the complex has homology $k \oplus k$ in degree 1 and k in degree 0. (If $k = \mathbf{C}$, then the corresponding variety X is homomorphic to a 2-torus minus one point, which is homotopic to a bouquet of two circles.)

Exercise 16.16: If S is any R-algebra and $F, G : S \to S$ are two R-linear derivations, then the commutator

$$[F, G] := FG - GF$$

is again a derivation. By the universal property of $d : S \to \Omega_{S/R}$, the maps F, G and $FG - GF$ must be of the form fd, gd, and hd, respectively, for some S-module homomorphisms $f, g, h : \Omega_{S/R} \to S$. Clearly,

$$FG - GF = fdgd - gdfd = (fdg - gdf)d,$$

so one might at first think that $h = (fdg - gdf)$. But this is wrong, since the right-hand side is not a homomorphism of S-modules! Show from the formula above that there is a homomorphism $k : \Omega_{S/R} \to S$ of R-modules such that $kd = 0$ and $h = fdg - gdf + k$. Of course, the formula $kd = 0$ suggests that k factors through the map defined above

$$\Omega_{B/A} = \Omega^1_{B/A} \xrightarrow{d^1} \Omega^2_{B/A}.$$

Show that this is correct by finding a map $k' : \Omega^2_{S/R} \to S$ of S-modules, depending on f and g, such that

$$h = fdg - gdf + k'd^1.$$

Exercise 16.17:* Proposition 16.12 gives a necessary and sufficient condition for the left-hand map in the conormal sequence to be a split monomorphism, but it may be a monomorphism without this condition being satisfied. The most important case is that of a radical complete intersection. Here is the result:

Let $(S, \mathfrak{m})$ be a local ring that is the localization of a finitely generated algebra over a field k, and let I be an ideal of S such that I/I^2 is a free module over S/I of rank equal to the codimension of I in S. (This is so whenever I is generated by a regular sequence, though not in very many other cases: See Exercises 17.16 and 20.23.) Set $T = S/I$ and suppose that, for each minimal prime P of T, the field $K(T/P)$ is separable over k and the ring S_P is regular. Consider the conormal sequence

$$I/I^2 \xrightarrow{d} T \otimes_S \Omega_{S/k} \to \Omega_{T/k} \to 0.$$

a. If I is a radical ideal, show that d is an injection.

b. One might hope that d would be an injection for all complete intersections. To dispel this hope, compute the kernel of d in the case $I = (x^2, y^2) \subset S = k[x, y]_{(x,y)}$.

Part III

Homological Methods

A **complex** of modules is a sequence of modules and maps between them

$$\mathcal{C}: \; \cdots \to A \xrightarrow{f} B \xrightarrow{g} C \to \cdots$$

such that the composition gf of any two consecutive maps is 0. Many of the operations that apply to modules, such as $\otimes$ and Hom, can also be applied to complexes of modules, as we shall see. Modules themselves may be regarded as complexes (where all the maps and all but one of the modules are 0). There is also a fundamental new operation, taking **homology**: The homology of the complex $\mathcal{C}$ at the module B is by definition the module $\ker(g)/\operatorname{im}(f)$.

Roughly speaking, using **homological methods** means studying modules and complexes of modules. These methods play something of the same role in commutative algebra that representation theory plays in the study of groups; although they seem at first rather roundabout, they are uncannily effective in solving certain problems. We have already seen an early application: Hilbert's use of free resolutions to compute Hilbert functions. Hilbert also wrote down the **Koszul complexes**, described in Chapter 17, as examples.

Perhaps the first serious application of homological ideas (and in fact of the Koszul complexes) was made by Arthur Cayley, who used them in elimination theory in [1858].[1] The first items we shall study, the Koszul complex and characterizations of its exactness, are closely related to Cayley's ideas.

Advice to the Reader: The chapters to come use more homological algebra (mostly the functors Tor and Ext) than the earlier chapters of this book. In addition, they make occasional use of the exterior algebra. The necessary topics—and much more besides—are treated in a compact way in Appendices A2 and A3, where there are also references to more leisurely treatments.

---

[1]Cayley regarded a complex in our sense as a parametrized family of complexes of vector spaces. The switch from polynomial as function to polynomial as element of a ring came only later, in the work of Kronecker.

17

Regular Sequences and the Koszul Complex

Throughout this chapter we shall assume that the rings considered are Noetherian. It is possible to do business without this condition, but some definitions should be changed slightly. The interested reader may consult Northcott [1976].

We recall a fundamental definition that extends the notion of a nonzerodivisor:

Definition. *Let R be a ring and let M be an R-module. A sequence of elements $x_1, \ldots, x_n \in R$ is called a* **regular sequence** *on M (or an M-* **sequence**) *if*

1. $(x_1, \ldots, x_n)M \neq M$, *and*

2. *For $i = 1, \ldots, n$, x_i is a nonzerodivisor on $M/(x_1, \ldots, x_{i-1})M$.*

Following ideas introduced by Auslander, Buchsbaum, and Serre in the 1950s and refined by many people since, we shall study this notion with a homological tool called the Koszul complex. (One can also go quite far without homological methods, buying an elementary treatment at the expense of a certain ingenuity—see Kaplansky [1970].) Most of this chapter is devoted to establishing the techniques for dealing with Koszul complexes. To make the content of the theory clear we shall start naively with the simplest cases and some simple applications; we then redo the theory from a more powerful point of view, using the exterior algebra.

We shall use the Koszul complex in this chapter to establish the basic facts about the notion of **depth**, an algebraic notion parallel to the geometric notion of codimension. In the next chapter we shall study the **Cohen-Macaulay** condition, which is the condition that the two notions coincide.

At the end of the chapter we present some common alternative descriptions of the Koszul complex, and explain the relation of the Koszul complex to the tangent and cotangent bundles of projective spaces, a relation that is the source of many beautiful applications. Some other geometric connections are given in the exercises.

The first part of the exposition is adapted from that of Buchsbaum [1969], which was my own introduction to this subject.

17.1 Koszul Complexes of Lengths 1 and 2

We can decide whether an element $x \in R$ is a nonzerodivisor from the homology of the complex

$$K(x) : 0 \to R \xrightarrow{x} R,$$

which is $(0 : x)$. This trivial remark is the essential basis for the homological study of regular sequences.

Given a second element $y \in R$, multiplication by y defines a map of complexes $K(x) \to K(x)$ — that is, a commutative diagram

$$\begin{array}{ccccc} K(x): & 0 & \to & R & \xrightarrow{x} & R \\ & & & \downarrow y & & \downarrow y \\ K(x): & 0 & \to & R & \xrightarrow{x} & R \end{array}$$

We can use the commutativity of the square in the diagram to build a larger complex, which we write schematically as

$$(*) \qquad K(x,y): \quad \begin{array}{ccccccccc} 0 & \to & R & \xrightarrow{x} & R & \to & 0 & & \\ & & & \searrow^{y} \oplus & & \searrow^{y} & & & \\ & & 0 & \to & R & \xrightarrow[-x]{} & R & \to & 0 \end{array}$$

or in more usual notation as

$$(**) \qquad K(x,y): \quad 0 \to R \xrightarrow{\binom{y}{x}} R \oplus R \xrightarrow{(-x,y)} R.$$

The reader who has seen Koszul complexes before may be shocked at our sign convention. The Koszul complex is usually written with signs

$$0 \to R \xrightarrow{\binom{y}{-x}} R \oplus R \xrightarrow{(x,y)} R.$$

These two possibilities are in fact isomorphic as complexes, so no harm has been done. We shall introduce the complex that is naturally written with

signs as in the second diagram in the section "Duality and Homotopies," at the end of this chapter; we shall write it as $K'(\varphi)$, where $\varphi : R^2 \to R$ is the map with matrix (x, y). The choice we are making has the advantage that the formula for the differential (in terms of exterior algebra) is more transparent.

For convenience, we shall number the homology groups of $K(x)$ and $K(x, y)$ starting from the left, and in deference to tradition we shall write these as cohomology groups; thus $H^0(K(x))$ is the homology at the leftmost nonzero term R of $K(x)$. We see from the definition that $H^0(K(x)) = (0 : x)$, the annihilator of x, and $H^0(K(x, y))$ is $(0 : (x, y))$, so that if x is a nonzerodivisor then $H^0(K(x, y)) = 0$.

What is the meaning of $H^1(K(x, y))$? First we analyze the kernel of the right-hand map. An element $(a, b) \in R \oplus R$ is in the kernel iff $-xa + yb = 0$. Of course this requires $b \in (x : y)$. Conversely, if $b \in (x : y)$, then there is an element a with $-xa + yb = 0$, so that (a, b) will be in the kernel. *If we assume that x is a nonzerodivisor*, then a is uniquely determined by b, and the association $b \mapsto a$ is a module homomorphism, so the kernel is isomorphic to $(x : y)$.

On the other hand, an element is in the image of the left-hand map iff it is of the form (cy, cx), so the elements of $(x : y)$ that correspond to elements of the image are the elements of (x). Thus if x is a nonzerodivisor, then

$$H^1(K(x, y)) \cong (x : y)/(x).$$

In particular, if x is a nonzerodivisor then $H^1(K(x, y)) = 0$ iff the sequence x, y satisfies condition 2 in the definition of a regular sequence.

A further point of interest can be deduced from the schematic presentation $(*)$ of $K(x, y)$: The lower row of this complex is actually a subcomplex (that is, it is taken into itself by the differential) isomorphic to $K(x)$, while the upper row, also isomorphic to $K(x)$, is the quotient of $K(x, y)$ by the lower row. As written, the $(i - 1)$ term of the subcomplex $K(x)$ is included in the ith term of $K(x, y)$, which projects to the ith term of the quotient $K(x)$, so the long exact sequence in homology coming from this short exact sequence of complexes has the form:

$$\cdots \to H^0(K(x)) \xrightarrow{\delta} H^0(K(x)) \to H^1(K(x, y)) \to H^1(K(x)) \to \cdots$$

where the map δ is the "connecting homomorphism." An easy diagram chase, which the reader should do, shows that δ is multiplication by y.

Now suppose only that $H^1(K(x, y)) = 0$. It follows from the preceding long exact sequence that

$$H^0(K(x))/yH^0(K(x)) = 0.$$

In general, not much can be deduced from this; but if we assume in addition that R is a Noetherian local ring and y is in the maximal ideal, then

Nakayama's lemma shows that $H^0(K(x)) = 0$! Consequently, x is a nonzerodivisor, and x, y is a regular sequence by what we have already proved. We may state what we have shown as follows:

Theorem 17.1. *If R is a Noetherian local ring and x, y are in the maximal ideal, then x, y is a regular sequence iff $H^1(K(x, y)) = 0$.*

From the way the Koszul complex is written in (**), it is clear that the complexes $K(x, y)$ and $K(y, x)$ are isomorphic. Thus under the hypothesis of Theorem 17.1, x, y is a regular sequence iff y, x is. This is enough to show that regular sequences may be permuted.

Corollary 17.2. *If R is a Noetherian local ring and $x_1, \ldots, x_r$ is a regular sequence of elements in the maximal ideal of R, then any permutation of $x_1, \ldots, x_r$ is again a regular sequence.*

Proof. Since every permutation is a product of transpositions of neighboring elements, it suffices to show that we can interchange two neighbors; that is, if $x_1, \ldots, x_i, x_{i+1}, \ldots, x_r$ is a regular sequence, then $x_1, \ldots, x_{i+1}, x_i, \ldots, x_r$ is too. The only part of the definition of a regular sequence that is not immediate for $x_1, \ldots, x_{i+1}, x_i, \ldots, x_r$ amounts to saying that x_{i+1}, x_i is a regular sequence modulo $(x_1, \ldots, x_{i-1})$. After factoring out $(x_1, \ldots, x_{i-1})$, Theorem 17.1 and the remark following it give the desired conclusion.

One might at first hope that the local hypothesis in these two results would be superfluous, required only by the clumsy methods used in the proof. This is not the case.

Example 17.3. Consider the ring

$$R = k[x, y, z]/(x - 1)z$$

and the sequence of elements

$$x, (x - 1)y.$$

The ideal they generate is $(x, (x - 1)y) = (x, y) \neq R$. Further, it is easy to see that x is a nonzerodivisor in R, and $R/(x) = k[y, z]/(z)$. Thus $x, (x-1)y$ is a regular sequence and

$$H^1(K(x, (x - 1)y)) = 0.$$

However, $(x - 1)y$ is a zerodivisor—it is killed by z—so that the sequence in reversed order is not a regular sequence. (Despite such examples, every ideal generated by a regular sequence is actually generated by a set of elements that is a regular sequence in any order; see Exercise 17.6.)

One further point is worth noting: If $x \in R$ is arbitrary, then $H^0(K(x, 0)) = H^0(K(x))$ (since both are isomorphic to $(0 : (x))$) even

though the complex $K(x, 0)$ is not isomorphic to $K(x)$. The less experienced reader may find it useful to stop and try at least Exercise 17.1 before proceeding.

17.2 Koszul Complexes in General

We could build up the Koszul complex step by step, iterating the process just illustrated (and we shall soon prove that this gives the correct answer), but the following construction is so direct, simple, and invariant that it has many advantages.

If N is any R-module, then the exterior algebra $\wedge N$ may be defined as the free algebra $R \oplus N \oplus (N \otimes N) \oplus \cdots$ modulo the relations $x \otimes y = -y \otimes x$ and $x \otimes x = 0$ for all x and y in N. The product of two elements a, b in $\wedge N$ will be written $a \wedge b$. $\wedge N$ is a graded algebra—the part of degree m, written $\wedge^m N$, is generated as an R-module by the products of exactly m elements of N. It is **skew commutative** in the sense that if a and b are homogeneous elements, then

$$a \wedge b = (-1)^{(\deg a)(\deg b)} b \wedge a,$$

and if a has degree 1, then $a \wedge a = 0$. (These two conditions are equivalent if 2 is a unit in R.) To avoid needing a notation for the degrees of elements, we shall usually "abuse notation" and write $(-1)^{ab}$ for $(-1)^{(\deg a)(\deg b)}$. Note that for any N we have $\wedge^0 N = R$.

The construction $\wedge N$ is functorial in N: That is, if $f : N \to M$ is a map of modules, then $\wedge f : \wedge N \to \wedge M$ is the map of algebras taking $a \wedge b \wedge \cdots$ to $fa \wedge fb \wedge \cdots$. If N is a free module (the only case we shall actually use) then the construction behaves just like the more familiar version where R is a field and N is a vector space. In particular, if N is free of rank n, then $\wedge^n N \cong R$, and if $f : N \to N$ is a map, then $\wedge^n f$ is multiplication by the determinant of any matrix representing f. In this case $\wedge^m N = 0$ for $m > n$.

Now given a module N and an element $x \in N$, we define the Koszul complex to be the complex

$$K(x) : 0 \to R \to N \to \wedge^2 N \to \cdots \to \wedge^i N \xrightarrow{d_x} \wedge^{i+1} N \to \cdots$$

where d_x sends an element a to the element $x \wedge a$; in particular, $1 \in R$ is sent to $d_x(1) = x \in N$. If N is free of rank n and

$$x = (x_1, \ldots, x_n) \in R^n \cong N,$$

then we shall sometimes write $K(x_1, \ldots, x_n)$ in place of $K(x)$.

One advantage of this definition is that it makes the functoriality of the Koszul complex obvious: If $f : N \to M$ is a map of modules sending $x \in N$ to $y \in M$, then the map $\wedge f : \wedge N \to \wedge M$ preserves the differential, and is thus a map of complexes, exactly because it is a map of algebras.

To gain familiarity with the Koszul complex, and because it will be important later, let us show that $H^n(K(x_1,\ldots,x_n)) = R/(x_1,\ldots,x_n)$. Set $N = R^n$, and consider the right-hand end of the Koszul complex:

$$\cdots \to \wedge^{n-1}N \to \wedge^n N \, \wedge^{n+1} N = 0.$$

Let $e_1,\ldots,e_n$ be a basis for $N = R^n$. We have $\wedge^n N \cong R$ by an isomorphism sending $e_1 \wedge \cdots \wedge e_n$ to 1. Similarly $\wedge^{n-1}N \cong R^n$, with basis $e_1 \wedge \cdots \wedge e_{i-1} \wedge \hat{e}_i \wedge e_{i+1} \wedge \cdots \wedge e_n$, for $i = 1,\ldots,n$, where the symbol $\hat{e}_i$ indicates that e_i has been omitted. Now the image of $e_1 \wedge \cdots \wedge e_{i-1} \wedge \hat{e}_i \wedge e_{i+1} \wedge \cdots \wedge e_n$ under the differential of the Koszul complex is

$$\left(\sum x_i e_i\right) \wedge e_1 \wedge \cdots \wedge e_{i-1} \wedge \hat{e}_i \wedge e_{i+1} \wedge \cdots \wedge e_n = \pm x_i e_1 \wedge \cdots \wedge e_n,$$

so the cokernel of $\wedge^{n-1}N \to \wedge^n N$ is isomorphic to $R/(x_1,\ldots,x_n)$.

In general, as suggested by the case of a Koszul complex of length 2, the homology of the Koszul complex $K(x_1,\ldots,x_n)$ has to do with regular sequences. It does not in general detect whether $x_1,\ldots,x_n$ is a regular sequence, but it detects something even more interesting: the lengths of the maximal regular sequences in the ideal $(x_1,\ldots,x_n)$. The result also shows that these lengths are all the same.

Theorem 17.4. *Let M be a finitely generated module over the ring R. If*

$$H^j(M \otimes K(x_1,\ldots,x_n)) = 0 \quad \textit{for } j < r$$

while

$$H^r(M \otimes K(x_1,\ldots,x_n)) \neq 0,$$

then every maximal M-sequence in $I = (x_1,\ldots,x_n) \subset R$ has length r.

We put off the proof until later in this chapter.

Corollary 17.5. *If $x_1,\ldots,x_n$ is an M-sequence, then $M \otimes K(x_1,\ldots,x_n)$ is exact except at the extreme right; that is, $H^j(M \otimes K(x_1,\ldots,x_n)) = 0$ for $j < n$. Furthermore, $H^n(M \otimes K(x_1,\ldots,x_n)) = M/(x_1,\ldots,x_n)M$.*

Proof. The length of a maximal M-sequence in $(x_1,\ldots,x_n)$ is clearly $\geq n$. The first conclusion follows from Theorem 17.4. For the second statement, writing N for R^n, we note that $H^n(M \otimes K(x_1,\ldots,x_n))$ is the homology of M tensored with the Koszul complex $M \otimes \wedge^{n-1}N \to M \otimes \wedge^n N \to 0$ at $M \otimes \wedge^n N$; that is, it is the cokernel of $M \otimes \wedge^{n-1}N \to M \otimes \wedge^n N$. By the right-exactness of the tensor product,

$$\begin{aligned}\operatorname{coker}(M \otimes \wedge^{n-1}N \to M \otimes \wedge^n N) &= M \otimes \operatorname{coker}(\wedge^{n-1}N \to \wedge^n N)\\ &= M \otimes H^n(K(x_1,\ldots,x_n)).\end{aligned}$$

Using the computation we already made of $H^n(K(x_1,\ldots,x_n))$ we see that

$$\begin{aligned}H^n(M \otimes K(x_1,\ldots,x_n)) &= M \otimes R/(x_1,\ldots,x_n)\\ &= M/(x_1,\ldots,x_n)M.\end{aligned}$$

□

That the converse is false may be seen from Example 17.3; but we shall see in Theorem 17.6 that the converse does hold in the local setting. We shall prove a more general version of this corollary in Corollary 17.12.

Note that if $IM \neq M$, then at least

$$H^n(M \otimes K(x_1, \ldots, x_n)) = M/(x_1, \ldots, x_n)M \neq 0,$$

while, of course,

$$H^{-1}(M \otimes K(x_1, \ldots, x_n)) = 0,$$

so there is an r for which Theorem 17.4 may be applied. On the other hand, we shall see that if $IM = M$, then $H^j(M \otimes K(x_1, \ldots, x_n)) = 0$ for every j.

If $IM \neq M$, then by Theorem 17.4 the lengths of all maximal M-sequences in I are the same. We define the depth of I on M, written **depth** $(\boldsymbol{I}, \boldsymbol{M})$, to be the length of any maximal M-sequence in I. If $M = R$, we shall speak simply of the depth of I. If $IM = M$, we adopt the convention that depth $(I, M) = \infty$.

When R is local with maximal ideal P, and M is an R-module, then we shall see in Chapter 19 that the depth of P on M, simply called the depth of M, is a particularly interesting number. This terminology conflicts with the one just introduced in case M is an ideal; however, confusion does not really arise in practice, and both pieces of terminology are commonly in use side by side. Some authors follow the original paper of Rees [1957] and resolve the problem by using the term "grade I" to denote the depth of I on R. But most geometers, following French usage, have adopted the term "depth," and we shall follow this tradition.

As we have mentioned, the depth of I is a kind of arithmetic measure of the size of I, while the codimension of I is a geometric measure. In Corollary 17.8 we shall show that, like the codimension, the depth depends only on the radical of I, and is thus geometric in the sense that it is, in the case of affine rings, say, determined by the variety cut out by I. Pursuing the depth–codimension analogy, Theorem 17.4 implies an analogue of the principal ideal theorem: Namely, it implies that an ideal with r generators can have depth at most r. We shall see that in general depth $I \leq \operatorname{codim} I$.

In the local case we can strengthen Theorem 17.4 further; it is enough for a single homology group to vanish, and we even get a criterion for a particular sequence to be an M-sequence:

Theorem 17.6. *Let M be a finitely generated module over the local ring $R, \mathfrak{m}$. Suppose $x_1, \ldots, x_n \in \mathfrak{m}$. If for some k*

$$H^k(M \otimes K(x_1, \ldots, x_n)) = 0,$$

then

$$H^j(M \otimes K(x_1, \ldots, x_n)) = 0 \quad \textit{for all } j \leq k.$$

In particular, if $H^{n-1}(M \otimes K(x_1, \ldots, x_n)) = 0$, then $x_1, \ldots, x_n$ is an M-sequence.

We shall postpone the proofs of Theorems 17.4 and 17.6 until we have developed some tools for handling Koszul complexes.

To avoid endlessly repeating the hypothesis, we shall use the letter M to denote a finitely generated R-module throughout the remainder of this chapter.

An immediate consequence of Theorem 17.6 strengthens Corollary 17.2.

Corollary 17.7. *If R is local and $(x_1, \ldots, x_n) \subset R$ is a proper ideal containing an M-sequence of length n, then $x_1, \ldots, x_n$ is an M-sequence.*

Proof. Since M is finitely generated, Nakayama's lemma shows that $H^n(M \otimes K(x_1, \ldots, x_n)) = M/(x_1, \ldots, x_n)M \neq 0$. If now r is the smallest number such that $H^r(M \otimes K(x_1, \ldots, x_n)) \neq 0$, then every maximal M-sequence in $(x_1, \ldots, x_n)$ has length r by Theorem 17.4. It follows from our hypothesis that $r = n$. Thus $x_1, \ldots, x_n$ is a regular sequence by Theorem 17.6.

This result can often be used to prove that a given sequence is regular. For example, we have:

Corollary 17.8 (Geometric nature of depth).

a. *If $x_1, \ldots, x_r$ is an M-sequence, then $x_1^t, \ldots, x_r^t$ is an M-sequence for any positive integer t.*

b. *Thus if I is an ideal of R and J is its radical, we have depth (I, M) = depth (J, M).*

Proof.

a. We do induction on r to reduce to the local case. For $r = 1$ we simply note that a power of a nonzerodivisor is a nonzerodivisor.

   For general r, we may suppose by induction that $x_1^t, \ldots, x_{r-1}^t$ is an M-sequence. It suffices to show that x_r is a nonzerodvisor on $M/(x_1^t, \ldots, x_{r-1}^t)M$ (that is, that multiplication by x_r has zero kernel). For this we may localize at a prime P. If P does not contain $x_1, \ldots, x_r$, then either $M/(x_1^t, \ldots, x_{r-1}^t)M = 0$ or x_r is a unit, and the result is obvious. Thus we may assume that R is a local ring and $x_1, \ldots, x_r$ is contained in the maximal ideal.

   If $x_1, \ldots, x_r$ is an M-sequence, then clearly $x_1, \ldots, x_{r-1}, x_r^t$ is an M-sequence. Since we are in a local ring, we may apply Corollary 17.7 and conclude that $x_r^t, x_1, \ldots, x_{r-1}$ is an M-sequence. Repeating the argument, we see that $x_{r-1}^t, x_r^t, x_1, \ldots, x_{r-2}$ is an M-sequence. After r such steps, we see that $x_1^t, \ldots, x_r^t$ is an M-sequence.

b. Since $I \subset J$ we have depth $(I, M) \leq$ depth (J, M) trivially. The opposite equality follows using part a since if $x_1, \ldots, x_r$ is an M-sequence in J, then $x_1^t, \ldots, x_r^t$ is in I for $t \gg 0$. □

17.3 Building the Koszul Complex from Parts

To analyze the Koszul complex, we use two methods of constructing complexes: tensor products and mapping cones. The reader new to these matters will find more details in the appendix on homological algebra. First, the **tensor product of two complexes**:

$$\mathcal{F} : \ \cdots \to F_i \xrightarrow{\varphi_i} F_{i+1} \to \cdots$$

and

$$\mathcal{G} : \ \cdots \to G_i \xrightarrow{\psi_i} G_{i+1} \to \cdots$$

is defined to be the complex

$$\mathcal{F} \otimes \mathcal{G} : \ \cdots \to \sum_{i+j=k} F_i \otimes G_j \xrightarrow{d_k} \sum_{i+j=k+1} F_i \otimes G_j \to \cdots$$

where the map d_k on $F_i \otimes G_j$ (with $i + j = k$) is the zero map to $F_s \otimes G_t$ unless $i = s$ or $j = t$, while from $F_i \otimes G_j$ to $F_{i+1} \otimes G_j$ it is $\varphi_i \otimes 1$ and from $F_i \otimes G_j$ to $F_i \otimes G_{j+1}$ it is $(-1)^i 1 \otimes \psi_j$. (The choice of sign is necessary—as the reader may easily check—to make $\mathcal{F} \otimes \mathcal{G}$ a complex, that is, to make

$$d_{k+1} d_k = 0.$$

Heuristically, we may say that the sign occurs "because" we have commuted the element ψ, of degree -1, and an element of F_i, of degree i, introducing the sign $(-1)^{(-1)i} = (-1)^i$).

We regard the modules in a complex as indexed in a fixed way, and if $\mathcal{G}$ is a complex we write $\mathcal{G}[n]$ for the complex $\mathcal{G}$ shifted n steps: That is, if G_i is the module in the ith position in $\mathcal{G}$, then G_{n+i} is the module in the ith position of $\mathcal{G}[n]$. It is convenient to take the differential of $\mathcal{G}[n]$ to be the differential of $\mathcal{G}$ multiplied by $(-1)^n$. Thus, for example, if we regard R as a complex $0 \to R \to 0$ with R in the zeroeth position, then $R[-i]$ will denote the complex $0 \to R \to 0$ with R in the ith position. Note that $\mathcal{G}[n] = R[n] \otimes \mathcal{G}$.

We now return to the Koszul complex. If $\mathcal{F}$ is the Koszul complex on one element $y \in R$,

$$\mathcal{F} : 0 \to R \xrightarrow{y} R \to 0,$$

then the obvious diagram

$$\begin{array}{lccccccc}
R[-1]: & 0 & \to & 0 & \to & R & \to & 0 \\
 & & & \downarrow & & \downarrow{\scriptstyle 1} & & \\
\mathcal{F}: & 0 & \to & R & \xrightarrow{y} & R & \to & 0 \\
 & & & \downarrow{\scriptstyle 1} & & \downarrow & & \\
R[0]: & 0 & \to & R & \to & 0 & \to & 0
\end{array}$$

is a short exact sequence of complexes

$$0 \to R[-1] \to \mathcal{F} \to R[0] \to 0.$$

If we tensor this diagram with another complex $\mathcal{G}$, then we get a short exact sequence of complexes $0 \to \mathcal{G}[-1] \to \mathcal{F} \otimes \mathcal{G} \to \mathcal{G} \to 0$. Indeed, $\mathcal{F} \otimes \mathcal{G}$ is the so-called **mapping cone** of the map $\mathcal{G}[-1] \to \mathcal{G}$ of complexes given by multiplication by y—that is, schematically as before, $\mathcal{F} \otimes \mathcal{G}$ is given by

$$\begin{array}{lccccccccc}
 & \cdots \to & G_i & \xrightarrow{-\varphi_i} & G_{i+1} & \xrightarrow{-\varphi_{i+1}} & G_{i+2} & \to \cdots \\
\mathcal{F} \otimes \mathcal{G}: & & \oplus & \searrow^{y} & \oplus & \searrow^{y} & \oplus & \\
 & \cdots \to & G_{i-1} & \xrightarrow[-\varphi_{i-1}]{} & G_i & \xrightarrow[-\varphi_i]{} & G_{i+1} & \to \cdots
\end{array}$$

Since $H^i(\mathcal{G}[-1]) = H^{i-1}(\mathcal{G})$, the short exact sequence of complexes gives rise to a long exact sequence in homology

$$\cdots \to H^{i-1}(\mathcal{G}) \xrightarrow{y} H^{i-1}(\mathcal{G}) \to H^i(\mathcal{F} \otimes \mathcal{G}) \to H^i(\mathcal{G}) \xrightarrow{y} \cdots,$$

where one checks by a direct diagram chase that the connecting homomorphisms are multiplication by y as claimed.

Proposition 17.9. *If $N = N' \oplus N''$, then $\wedge N = \wedge N' \otimes \wedge N''$ as skew-commutative algebras. If $x' \in N'$ and $x'' \in N''$ are elements, so that $x = (x', x'') \in N$, then*

$$K(x) = K(x') \otimes K(x'')$$

as complexes.

Proof. The first statement of the proposition is proven in Proposition A2.2 of Appendix 2.

To prove the second statement, it suffices to note that if $y = y' \otimes y'' \in \wedge N' \otimes \wedge N'' = \wedge N$, and

$$\begin{aligned}
x = (x', x'') = x' \otimes 1 + 1 \otimes x'' &\in \wedge^1 N' \otimes R \oplus R \otimes \wedge^1 N'' \\
&= \wedge^1 N' \otimes \wedge^0 N'' \oplus \wedge^0 N' \otimes \wedge^1 N'' \\
&= \wedge^1 N,
\end{aligned}$$

then

$$\begin{aligned}
x \wedge y &= (x' \otimes 1 + 1 \otimes x'') \wedge (y' \otimes y'') \\
&= (x' \wedge y') \otimes y'' + (-1)^{x''y'} y' \otimes (x'' \wedge y''),
\end{aligned}$$

so the differentials in $\wedge N$ and $\wedge N' \otimes \wedge N''$ agree under the identification above.

We shall prove Theorem 17.4 by applying Proposition 17.9 in two ways. The first shows how the Koszul complex of $x_1, \ldots, x_n$ can reflect information about regular sequences contained in the ideal generated by $x_1, \ldots, x_n$.

Corollary 17.10. *If $y_1, \ldots, y_r$ are elements of the ideal generated by $x_1, \ldots, x_n \in R$, and M is any R-module, then*

$$\begin{aligned} &H^*(M \otimes K(x_1, \ldots, x_n, y_1, \ldots, y_r)) \\ &\quad \cong H^*(M \otimes K(x_1, \ldots, x_n)) \otimes \wedge R^r \end{aligned}$$

as graded modules, In particular, for each i we have

$$\begin{aligned} &H^i(M \otimes K(x_1, \ldots, x_n, y_1, \ldots, y_r)) \\ &\quad \cong \sum_{i=j+k} H^k(M \otimes K(x_1, \ldots, x_n)) \otimes \wedge^j R^r. \end{aligned}$$

Thus,

$$H^i(M \otimes K(x_1, \ldots, x_n, y_1, \ldots, y_r)) = 0$$

iff

$$H^k(M \otimes K(x_1, \ldots, x_n)) = 0 \quad \textit{for all } k \textit{ with } i - r \leq k \leq i.$$

Proof. There is an automorphism of $R^n \oplus R^r$ taking the element with coordinates $x_1, \ldots, x_n, y_1, \ldots, y_r$ to the one with coordinates $x_1, \ldots, x_n, 0, \ldots, 0$ (r zeros); indeed, if $y_i = \sum_j a_{ij} x_j$, and A is the $r \times n$ matrix with entries a_{ij}, then the matrix

$$\left(\begin{array}{c|c} I & 0 \\ -A & I \end{array} \right)$$

takes the column vector with entries $x_1, \ldots, x_n, y_1, \ldots, y_r$ to the one with entries $x_1, \ldots, x_n, 0, \ldots, 0$. From the functoriality of the Koszul complex and Proposition 17.9 we get

$$\begin{aligned} K(x_1, \ldots, x_n, y_1, \ldots, y_r) &\cong K(x_1, \ldots, x_n, 0, \ldots, 0), \\ &\cong K(x_1, \ldots, x_n) \otimes K(0, \ldots, 0). \end{aligned}$$

Now $K(0, \ldots, 0)$ is the exterior algebra on r generators, with differentials all 0, whence the first statement of the corollary. The last two statements follow immediately.

Applying Proposition 17.9 in the case $N = R \oplus N''$ and using the remarks just before the proposition, we get:

Corollary 17.11. *If $x = (x', y) \in N = N' \oplus R$, then $K(x)$ is isomorphic to the mapping cylinder of the map $K(x') \to K(x')$ induced by multiplication by y; in particular, we have a long exact sequence:*

$$\cdots \to H^i(M \otimes K(x')) \xrightarrow{y} H^i(M \otimes K(x')) \to H^{i+1}(M \otimes K(x)) \\ \to H^{i+1}(M \otimes K(x')) \xrightarrow{y} \cdots.$$

Proof. Since $N' \oplus R \cong R \oplus N'$ in a way taking (x', y) to (y, x'), we have $K(x) = K(y, x')$ by functoriality. By Proposition 17.9, $K(x) = K(y) \otimes K(x')$ and by the remark before the proposition such a tensor product is a mapping cylinder, so we have a short exact sequence of complexes

$$0 \to M \otimes K(x')[-1] \to M \otimes K(x) \to M \otimes K(x') \to 0,$$

where we have written

$$K(x')[-1]$$

for the complex whose ith term is the $(i-1)$ term of $K(x')$. The desired long exact sequence is just the long exact sequence in homology of this short exact sequence of complexes.

From Corollary 17.11 we obtain a more precise version of part of Theorem 17.4.

Corollary 17.12. *If $x_1, \ldots, x_i$ is an M-sequence, then*

$$H^i(M \otimes K(x_1, \ldots, x_n)) \\ = ((x_1, \ldots, x_i)M : (x_1, \ldots, x_n))/(x_1, \ldots, x_i)M.$$

In particular, if $x_1, \ldots, x_i$ is an M-sequence in the ideal $I = (x_1, \ldots, x_n)$, then $H^j(M \otimes K(x_1, \ldots, x_n)) = 0$ for $j < i$. If $x_1, \ldots, x_i$ is a maximal M-sequence in I, and $IM \neq M$, then $H^i(M \otimes K(x_1, \ldots, x_n)) \neq 0$.

Proof. We prove the first statement by induction on i, starting with $i = 0$, where the statement follows directly from the definition of the Koszul complex.

For given i, we do induction on n, starting from $n = i$. If $n = i$ then the first statement becomes $H^i(M \otimes K(x_1, \ldots, x_n)) = M/(x_1, \ldots, x_n)M$, which is clear from the definition of the Koszul complex. Now suppose $n > i$. By the induction on i we have

$$H^{i-1}(M \otimes K(x_1, \ldots, x_n)) \\ = ((x_1, \ldots, x_{i-1})M : (x_1, \ldots, x_n))/(x_1, \ldots, x_{i-1})M \\ = 0$$

since x_i is a nonzerodivisor on $M/(x_1, \ldots, x_{i-1})M$. Thus the exact sequence of Corollary 17.11 yields

$$H^i(M \otimes K(x_1, \ldots, x_n)) = \\ \ker(H^i(M \otimes K(x_1, \ldots, x_{n-1})) \xrightarrow{x_n} H^i(M \otimes K(x_1, \ldots, x_{n-1})).$$

Since also

$$((x_1,\ldots,x_i)M:(x_1,\ldots,x_n))/(x_1,\ldots,x_i)M =$$
$$\ker(((x_1,\ldots,x_i)M:(x_1,\ldots,x_{n-1}))/(x_1,\ldots,x_i)M \xrightarrow{x_n}$$
$$((x_1,\ldots,x_i)M:(x_1,\ldots,x_{n-1}))/(x_1,\ldots,x_i)M),$$

we are done.

The vanishing part of the second statement follows from the first since if x_{j+1} is a nonzerodivisor on $M/(x_1,\ldots,x_j)M$ then

$$((x_1,\ldots,x_j)M:(x_1,\ldots,x_n)) = (x_1,\ldots,x_j)M.$$

To deduce the nonvanishing part, suppose that $(x_1,\ldots,x_i)$ is a maximal M-sequence in I. This means that I is contained in the set of zerodivisors on $M/(x_1,\ldots,x_i)M$. By the theory of associated primes this set of zerodivisors is a finite union of associated primes, so by prime avoidance, Lemma 3.3, I must be contained in a single associated prime Q of $M/(x_1,\ldots,x_i)M$. By definition, Q is the annihilator of some nonzero element m of $M/(x_1,\ldots,x_i)M$, so we see that

$$m \in ((x_1,\ldots,x_i)M:(x_1,\ldots,x_n))/(x_1,\ldots,x_i)M,$$

whence the nonvanishing statement.

We turn to the proofs of Theorems 17.4 and 17.6.

Proof of Theorem 17.4. Let $y_1,\ldots,y_s$ be a maximal M-sequence in I. By hypothesis, r is the smallest integer i such that

$$H^i(M\otimes K(x_1,\ldots,x_n)) \neq 0.$$

By Corollary 17.10, r is also the smallest for which

$$H^i(M\otimes K(x_1,\ldots,x_n,y_1,\ldots,y_s)) \neq 0.$$

From the hypothesis $H^r(M \otimes K(x_1,\ldots,x_r)) \neq$ it follows that $(x_1,\ldots,x_r)M \neq M$; this will be proved in Proposition 17.4, below. Using Corollary 17.12 we get $s = r$, as required.

Proof of Theorem 17.6. We prove the first statement by induction on n. If $H^k(M\otimes K(x_1,\ldots,x_n)) = 0$, then by Corollary 17.11 the map

$$H^{k-1}(M\otimes K(x_1,\ldots,x_{n-1})) \xrightarrow{x_n} H^{k-1}(M\otimes K(x_1,\ldots,x_{n-1}))$$

is an epimorphism. By Nakayama's lemma, this implies that

$$H^{k-1}(M\otimes K(x_1,\ldots,x_{n-1})) = 0,$$

so by induction,

$$H^j(M\otimes K(x_1,\ldots,x_{n-1})) = 0 \quad \text{for } j \leq k-1.$$

Using Corollary 17.11 again, we see that

$$H^j(M \otimes K(x_1, \ldots, x_n)) = 0 \quad \text{for } j \leq k$$

as required for the first statement.

To prove the second statement we use the same strategy. If

$$H^{n-1}(M \otimes K(x_1, \ldots, x_n)) = 0$$

then, as just noted,

$$H^{n-2}(M \otimes K(x_1, \ldots, x_{n-1})) = 0$$

so by induction $x_1, \ldots, x_{n-1}$ is an M-sequence. By Corollary 17.12,

$$\begin{aligned} 0 &= H^{n-1}(M \otimes K(x_1, \ldots, x_n)) \\ &= ((x_1, \ldots, x_{n-1})M : (x_1, \ldots, x_n))/(x_1, \ldots, x_{n-1})M, \end{aligned}$$

so x_n is a nonzerodivisor on $M/(x_1, \ldots, x_{n-1})M$, as required. □

17.4 Duality and Homotopies

There is a dual version of the Koszul complex, associated to an R-module N and a map $\varphi : N \to R$. The description requires a little exterior algebra. The inexperienced reader may wish to stick to the case where N is free, in which case what we shall describe is just the dual of the Koszul complex $K(\varphi)$, where $\varphi \in N^*$. Alternately, all the required algebra is done in detail in the Appendix 2.

Corresponding to $\varphi : N \to R$ we shall describe a complex

$$K'(\varphi) : \cdots \to \wedge^i N \xrightarrow{\delta_\varphi} \wedge^{i-1} N \to \cdots \to N \xrightarrow{\varphi} R \to 0.$$

To describe the differential $\delta_\varphi : \wedge^i N \to \wedge^{i-1} N$ we use the "diagonalization map" $\Delta : \wedge N \to \wedge N \otimes_R \wedge N$, the unique map of algebras taking each $m \in N = \wedge^1 N$ to

$$m \otimes 1 - 1 \otimes m \in \wedge^1 N \otimes_R \wedge^0 N \oplus \wedge^0 N \otimes_R \wedge^1 N \subset \wedge N \otimes_R \wedge N.$$

We shall actually use only the component of Δ that maps $\wedge^i N$ to $N \otimes (\wedge^{i-1} N)$, which may be described on generators as

$$\Delta'(m_1 \wedge \cdots \wedge m_i) = \sum_{j=1}^{i} (-1)^{j-1} m_j \otimes m_1 \wedge \cdots \wedge \hat{m}_j \wedge \cdots \wedge m_i,$$

where $\hat{m}_j$ signifies that m_j has been left out of the product. We define δ_φ to be the composite

$$\wedge^i N \xrightarrow{\Delta'} N \otimes_R (\wedge^{i-1} N) \xrightarrow{\varphi \otimes 1} R \otimes_R \wedge^{i-1} N = \wedge^{i-1} N.$$

When $i = 1$ this is nothing but φ.

To show we have defined a complex, we must show that $\delta_\varphi^2 = 0$. Computing, we see that $\delta_\varphi^2(n_1 \wedge \cdots \wedge n_i)$ is a linear combination of the terms $n_1 \wedge \cdots \wedge \hat{n}_j \wedge \cdots \wedge \hat{n}_{j'} \wedge \cdots \wedge n_i$, where we have left out n_j and $n_{j'}$. Assuming that $j < j'$ we see that the coefficient of this term in $\delta_\varphi^2(n_1 \wedge \cdots \wedge n_i)$ is

$$(-1)^j(-1)^{j'-1}\varphi(n_j)\varphi(n_{j'}) + (-1)^j(-1)^{j'}\varphi(n_j)\varphi(n_{j'}) = 0,$$

so indeed $\delta_\varphi^2 = 0$.

Many computations such as this one can be simplified by using Proposition A2.8 proved in the appendix on multilinear algebra, that δ_φ is a derivation of the exterior algebra $\wedge N$ to itself. That is, if n, n' are homogeneous elements of $\wedge N$, then

$$\delta_\varphi(n \wedge n') = \delta_\varphi(n) \wedge n' + (-1)^n n \wedge \delta_\varphi(n').$$

The sign is the natural one for a degree -1 derivation of skew-commutative graded algebras. To prove that $\delta_\varphi^2 = 0$ using this, we argue that $\delta_\varphi^2(m) = 0$ by induction on the degree of m. If $\deg m = 0$ the formula is obvious. In general, we may write m as a linear combination of elements of the form $n \wedge n'$, where n has degree 1. We have

$$\begin{aligned}
\delta_\varphi\delta_\varphi(n \wedge n') &= \delta_\varphi(\delta_\varphi(n) \wedge n') - \delta_\varphi(n \wedge \delta_\varphi(n')) \\
&= \delta_\varphi(\delta_\varphi(n)) \wedge n' + \delta_\varphi(n) \wedge \delta_\varphi(n') \\
&\qquad - \delta_\varphi(n) \wedge \delta_\varphi(n') + n \wedge \delta_\varphi\delta_\varphi(n') \\
&= \delta_\varphi(n) \wedge \delta_\varphi(n') - \delta_\varphi(n) \wedge \delta_\varphi(n') \\
&= 0.
\end{aligned}$$

Here is one relation between the Koszul complex defined originally and this dual version:

Lemma 17.13 (Homotopy for the Koszul complex). *If $x \in N$ and $\varphi : N \to R$, then the maps δ_φ and d_x satisfy the identity*

$$d_x\delta_\varphi + \delta_\varphi d_x = \varphi(x) \cdot 1,$$

where 1 is the identity map on $\wedge N$. Thus δ_φ is a homotopy showing that multiplication by $\varphi(x)$ is homotopic to 0 on $K(x)$, and similarly for d_x on $K'(\varphi)$.

Proof. The proof is a straightforward computation. Indeed, it is trivial if we use the fact that δ_φ is a derivation. We have

$$\begin{aligned}
d_x\delta_\varphi(n) + \delta_\varphi d_x(n) &= x \wedge \delta_\varphi(n) + \delta_\varphi(x \wedge n) \\
&= x \wedge \delta_\varphi(n) + \delta_\varphi(x) \wedge n - x \wedge \delta_\varphi(n) \\
&= \delta_\varphi(x) \wedge n \\
&= \varphi(x)n.
\end{aligned}$$

Of course, this can also be shown directly: We have

$$\begin{aligned}\delta_\varphi d_x(n_1 \wedge \cdots \wedge n_i) &= \delta_\varphi(x \wedge n_1 \wedge \cdots \wedge n_i)\\ &= \varphi(x) \cdot n_1 \wedge \cdots \wedge n_i\\ &\qquad + \sum_j (-1)^j x \wedge n_1 \wedge \cdots \wedge \hat{n}_j \wedge \cdots \wedge n_i;\\ d_x\delta_\varphi(n_1 \wedge \cdots \wedge n_i) &= x \wedge \sum_j (-1)^{j-1} n_1 \wedge \cdots \wedge \hat{n}_j \wedge \cdots \wedge n_i.\end{aligned}$$

If we add, all the terms cancel except $\varphi(x) \cdot n_1 \wedge \cdots \wedge n_i$.

For yet another proof, in a special case, see Exercise 17.7.

Here are some consequences:

Proposition 17.14.

a. *If $y \in (x_1, \ldots, x_n)$, then y annihilates the Koszul homology groups $H^j(M \otimes K(x_1, \ldots, x_n))$ for all M and all j.*

b. *If $(x_1, \ldots, x_n)M = M$, then $H^j(M \otimes K(x_1, \ldots, x_n)) = 0$ for all j.*

Proof.

a. If $y = \sum a_i x_i$, then the map $\varphi : R^n \to R$ with matrix $(a_1, \ldots, a_n)$ carries $x := (x_1, \ldots, x_n) \in R^n$ to $y \in R$. Thus, by the lemma, δ_φ is a homotopy showing that multiplication by $\varphi(x) = y$ on $K(x_1, \ldots, x_n)$ is homotopic to 0. Since a map homotopic to 0 induces the zero map on homology, we are done.

b. We may replace R by $R/(\text{annihilator } M)$ without changing $M \otimes K(x_1, \ldots, x_n)$, so we may assume that the annihilator of M is 0. By Corollary 4.7, we see that there is an element $y \in (x_1, \ldots, x_n)$ such that $1 - y$ annihilates M; thus $y = 1$. Now apply part a. □

If M is an R-module and $x \in M$ then x corresponds to a functional $x^* : M^* \to R$ given by $x^*(\varphi) = \varphi(x)$. Thus we may define Koszul complexes $K(x)$ and $K'(x^*)$. If M is a free module (and as always finitely generated), then an examination of the terms suggests that these complexes are dual to one another, and this is true. More surprisingly, they are isomorphic; that is, the Koszul complex is self-dual. We make the isomorphism explicit as follows.

As we have already noted, $\wedge^n R^n \cong R$. If we fix such an isomorphism $f : \wedge^n R^n \to R$ (called an "orientation" of R^n because if $R = \mathbf{R}$, the real numbers, then this corresponds exactly to the geometric notion of orientation), then there are induced isomorphisms $\wedge^k R^n \to \wedge^{n-k}(R^n)^* \to \wedge^{n-k}(R^{n*})$ that may be defined as follows: First, given elements $a \in \wedge^k R^n$ and $b \in \wedge^{n-k}(R^n)$, we define $\alpha : \wedge^k R^n \to \wedge^{n-k}(R^n)^*$ by setting $\alpha(a)(b) =$

$f(a \wedge b)$. The map α is an isomorphism because if $\{e_1, \ldots, e_n\} \subset R^n$ is a basis, then $\{\alpha(e_{i_1} \wedge \cdots \wedge e_{i_k})\}$ is, up to sign, a dual basis to $\{e_{j_1} \wedge \cdots \wedge e_{j_{n-k}}\}$.

Next we define $\beta : (\wedge^{n-k} R^n)^* \to \wedge^{n-k}(R^{n*})$. Suppose that $\{f_1, \ldots, f_n\} \subset R^{n*}$ is a basis dual to $\{e_1, \ldots, e_n\}$. β takes the basis element dual to $e_{i_1} \wedge \cdots \wedge e_{i_k}$ to the element $(-1)^s f_{j_1} \wedge \cdots \wedge f_{j_{n-k}}$, where $\{i_1, \ldots, i_k\} \cup \{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\}$ and $(-1)^s$ is the sign of the permutation $[i_1, \ldots, i_k, j_1, \ldots, j_{n-k}]$.

Using these isomorphisms we have:

Proposition 17.15 (Self-duality of the Koszul complex). *For $x \in R^n$ there is a commutative diagram*

$$\begin{array}{lcccc}
K(x): & \cdots \to & \wedge^i R^n & \xrightarrow{d_x} & \wedge^{i+1} R^n \to \cdots \\
 & & \downarrow \alpha & & \downarrow \alpha \\
K(x)^*: & \cdots \to & (\wedge^{n-i} R^n)^* & \xrightarrow{d_x^*} & (\wedge^{n-i-1} R^n)^* \to \cdots \\
 & & \uparrow \beta & & \uparrow \beta \\
K(x^*): & \cdots \to & \wedge^{n-i}(R^{n*}) & \xrightarrow{\delta_{x*}} & \wedge^{n-i-1}(R^{n*}) \to \cdots
\end{array}$$

where the vertical maps α and β are isomorphisms.

Since we shall not make direct use of this fact, we leave the proof to the reader in Exercise 17.8.

This explains at last our convention of writing the homology of the Koszul complex as cohomology: With the obvious notation, we have

$$H^k(K(x), d_x) = H_{n-k}(K(x), \delta_{x*}) = H^k(\mathrm{Hom}_R(K(x), R), d_x^*).$$

The two right-hand forms are the ones that usually occur in the literature, despite the advantage of simplicity in the form we have adopted.

17.5 The Koszul Complex and the Cotangent Bundle of Projective Space

A major geometric source of the significance of the Koszul complex is its relation to the tangent bundle on projective n-space. We shall only sketch the connection, which belongs perhaps more to algebraic geometry than to commutative algebra. First note that if R is a graded ring and $x = (x_0, \ldots, x_n)$ is a sequence of homogeneous elements, then the Koszul complex $K(x)$ has the structure of a graded complex (that is, a complex of graded free modules where all the maps are homogeneous of degree 0). Quite generally, if M is a graded R-module, then since $\wedge M$ is a bigraded algebra, the Koszul complex of any homogeneous element of degree 0 of M is graded. Thus we need only regard x as a homogeneous element of degree

0 in the graded free module $M = \oplus_i R(\deg x_i)$. If all the x_i have degree 1, then we can write the jth term of the complex as $\wedge^j R^{n+1}(j)$.

Now take R to be the graded polynomial ring $k[x_0, \ldots, x_n]$ and let T be the cokernel of the first map in the Koszul complex,

$$0 \to R \to R^{n+1}(1) \to T \to 0.$$

It is well known that T is the graded module corresponding to the (sheaf of algebraic sections of the) tangent bundle $\mathcal{T}$ of projective space under the usual Serre correspondence between sheaves on $\mathbf{P}^n$ and graded R-modules. (A part of this is obvious: Up to twist, the exact sequence above simply expresses the fact that, since $\mathbf{P}^n$ is the set of lines through the origin in the vector space k^{n+1}, the tangent space to $\mathbf{P}^n$ at a point corresponding to a line L is the quotient of the tangent space to k^{n+1} at a point of L modulo a tangent vector to L at that point.)

In geometry, not only the tangent bundle $\mathcal{T}_X$ of a smooth variety X, but also its exterior powers and those of its dual, the cotangent bundle, are interesting. For example, the exterior powers of the cotangent bundle

$$\Omega_X^q := \wedge^q \mathcal{T}_X^*$$

appear in the Hodge formula for singular cohomology over the complex numbers:

$$H^t(X, \mathbf{C}) = \oplus_{p+q=t} H^p(X, \Omega_X^q),$$

and the highest exterior power $\omega_X = \Omega^{\dim X}$ plays a central role in duality. It turns out that the Koszul complex connects the graded modules corresponding to these sheaves in case $X = \mathbf{P}^n$.

Theorem 17.16. *Let $R = k[x_0, \ldots, x_n]$ be a polynomial ring in $n+1$ variables over a field k, and let $K(x)$ be the Koszul complex of the sequence $x = (x_0, \ldots, x_n)$. Let T_j be the jth module of cycles in the Koszul complex, that is*

$$T_j = \ker\left(\wedge^{j+1} R^{n+1}(j+1) \to \wedge^{j+2} R^{n+1}(j+2)\right).$$

Then $T_j = \wedge^j T$ for $j = 0, \ldots, n-1$. Further, for all j, T_j is the graded module associated to the bundle $\wedge^j \mathcal{T}_{\mathbf{P}^n}$, and $T_{n-j} \cong \wedge^j T^*(n+1)$ is the module associated to the bundle $\Omega^j_{\mathbf{P}^n}(n+1)$.

Of course a similar result holds for the Koszul complex of any regular sequence.

Proof Sketch. The first statement follows from the right-exactness of the exterior algebra: If $F \to G \to M \to 0$ is any right-exact sequence of R-modules, then we get a right-exact sequence

$$F \otimes \wedge G \to \wedge G \to \wedge M \to 0,$$

and thus

$$\wedge^j M = \operatorname{coker} F \otimes \wedge^{j-1} G \to \wedge^j G,$$

the map being the multiplication in the exterior algebra. Applying this to T we see that $\wedge^j T$ is the cokernel of the Koszul complex map

$$R \otimes \wedge^{j-1} R^{n+1}(j-1) = \wedge^{j-1} R^{n+1}(j-1) \xrightarrow{d_x} \wedge^j R^{n+1}(j)$$

as required.

To simplify the notation let $X = \mathbf{P}^n$. The canonical bundle on X is $\Omega_X \cong \mathcal{O}_{\mathbf{P}^n}(-n-1)$. Equivalently, $\wedge^n \mathcal{T}_X = \mathcal{O}_{\mathbf{P}^n}(n+1)$, so the pairing

$$\wedge^j \mathcal{T}_X \otimes \wedge^{n-j} \mathcal{T}_X \to \wedge^n \mathcal{T}_X$$

gives us $\wedge^{n-j} \mathcal{T}_X \cong (\wedge^j \mathcal{T}_X^*)(n+1) = \Omega_X^j(n+1)$. Thus all the rest of the statements will follow once we show that T_j is the graded module associated to the bundle $\wedge^j \mathcal{T}_X$. For this it suffices to show that the depth of the module $T_j \geq 2$ (see the discussion of local cohomology in Appendix A4). Of course $T_n = R(n+1)$, while $T_j = 0$ for $j > n$, so these cases present no problem. In the interesting cases, a stronger result is given in Exercise 17.9. □

17.6 Exercises

We continue to assume that all rings considered are Noetherian.

Exercise 17.1: Suppose that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is a matrix over R whose determinant is a unit in R. Using the primitive description of the Koszul complex for two elements given at the beginning of the chapter, construct an isomorphism $K(x,y) \to K(ax+by, cx+dy)$. (A more general form of this follows trivially from our construction of the Koszul complex from the exterior algebra.) Conclude that if R is local and x, y are in the maximal ideal, then x, y is a regular sequence iff $ax+by, cx+dy$ is a regular sequence.

Exercise 17.2:* If $a \neq 0, b$ are elements of a domain R, then the element $ax - b \in R[x]$ is prime iff $H^1(K(a,b)) = 0$, in which case $R[x]/(ax-b) \cong R[b/a] \subset K(R)$, the quotient field of R. (Compare Exercise 10.4, and see Exercise 17.14b) for a generalization.)

Exercise 17.3: Here are easy parallel characterizations of regular sequences and systems of parameters (see Exercise 17.16 for a further analogy). Let $x_1, \ldots, x_n \in R$ be a sequence of elements in the maximal ideal of a local ring.

a. Show that $x_1, \ldots, x_n$ is a regular sequence iff for each $i = 0, \ldots, n-1$ the element x_{i+1} is not in any associated prime of $(x_1, \ldots, x_i)$.

b. Show that $x_1, \ldots, x_n$ is part of a system of parameters iff for each $i = 0, \ldots, n-1$ the element x_{i+1} is not in any minimal prime of $(x_1, \ldots, x_i)$.

Exercise 17.4 (Kaplansky): Abstract the method of Corollary 17.8 to show that if $x_1, \ldots, x_r$ is an M-sequence and $x_2, \ldots, x_r$ is an M-sequence, then $x_2, \ldots, x_r, x_1$ is an M-sequence.

Exercise 17.5: Show that if $x_1, \ldots, x_r$ is a regular sequence in R then so is $x_1^{a_1}, \ldots, x_r^{a_r}$ for any positive integers a_i. (See Exercise 17.13d for a much stronger result with slightly stronger hypotheses.)

Exercise 17.6:* Show that if an ideal I in a Noetherian ring R can be generated by a regular sequence, then it can be generated by a set of elements that is a regular sequence in any order (I learned this from Craig Huneke).

Exercise 17.7:

a. Suppose that $\alpha, \beta : \mathcal{F} \to \mathcal{F}'$ are maps of complexes, homotopic by a homotopy $\sigma : \mathcal{F} \to \mathcal{F}'[-1]$. If $\mathcal{G}$ is another complex, show that the maps $\alpha \otimes 1, \beta \otimes 1 : \mathcal{F} \otimes \mathcal{G} \to \mathcal{F}' \otimes \mathcal{G}$ are homotopic by a homotopy that may be described as $\sigma \otimes 1$.

b. If $x \in R$, show that multiplication by x is homotopic to 0 on $K(x)$.

c. If multiplication by each of $x_1, \ldots, x_n \in R$ is homotopic to 0 on a complex $\mathcal{F}$, show that for any $x \in (x_1, \ldots, x_n)$ the multiplication by x is homotopic to 0 on $\mathcal{F}$. Show that multiplication by x_i is homotopic to 0 on $K(x_1, \ldots, x_n)$ by showing that

$$K(x_1, \ldots, x_n) \cong K(x_1, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_n) \otimes K(x_i)$$

and using part a.

Exercise 17.8: Prove Proposition 17.15. Note that the identification of $\wedge^{n-k}(R^{n*})$ with $(\wedge^{n-k} R^n)^*$ requires some signs.

Exercise 17.9:* Complete the proof of Theorem 17.16 by proving that depth $T_j = n + 1 - j$ for $j = 0, \ldots, n$.

Exercise 17.10 (Koszul homology as Tor):

a. Suppose that $S \to R$ is a homomorphism of rings, and suppose that $y_1, \ldots, y_r \in S$ have images $x_1, \ldots, x_r$ in R. If $y_1, \ldots, y_r$ is a regular sequence in S, show that

$$H_i(K'(x_1, \ldots, x_r) \otimes M) = \operatorname{Tor}_i^S(S/(x_1, \ldots, x_r), M).$$

b. Show that for any ring R and any sequence of elements $x_1, \ldots, x_r \in R$ there is a ring S, a regular sequence $y_1, \ldots, y_r$ in S, and a ring homomorphism carrying y_i to x_i.

Free Resolutions of Monomial Ideals

Exercise 17.11 (Taylor's resolution of a monomial ideal):* It remains an open problem to give in closed form the minimal free resolution of an arbitrary monomial ideal. However, there is a nice nonminimal resolution that was discovered by Diana Taylor [1960]. It generalizes the Koszul complex in a natural way. Let $S = A[x_1, \ldots, x_r]$, where A is any ring and the x_i are indeterminates. Let $m_1, \ldots, m_t$ be monomials in the x_i. Define the **Taylor complex** $T(m_1, \ldots, m_t)$ as follows. Let F_s be the free module on basis elements e_I, where I is a subset $\{1, \ldots, t\}$. Set

$$m_I = \text{least common multiple } \{m_i | i \in I\}.$$

For each pair of subsets I, J such that I has s elements and J has $s-1$ elements, let $I = \{i_1, \ldots, i_s\}$ and suppose that $i_1 < \ldots < i_s$. Define:

$$c_{I,J} = \begin{cases} 0 \text{ if } J \not\subset I \\ (-1)^k m_I/m_J \text{ if } I = J \cup \{i_k\} & \text{for some } k. \end{cases}$$

Finally, define

$$d_s : F_s \to F_{s-1}$$

by sending e_I to $\sum_J c_{I,J}$. Let

$$T(m_1, \ldots, m_t) : 0 \to F_t \xrightarrow{d_t} \cdots \xrightarrow{d_1} F_0.$$

Show that $T(m_1, \ldots, m_t)$ is a free resolution as follows:

a. Given any monomial m, let n_i be a generator of $(m_i : m)$ (in other words $n_i = m_i/\mathrm{GCD}(m, m_i)$). Show that there is a map of complexes $\varphi_m : T(n_1, \ldots, n_t) \to T(m_1, \ldots, m_t)$ sending $e_\emptyset$ to $me_\emptyset$.

b. Show that (up to an adjustment of signs) the mapping cylinder of φ_m is the complex $T(m_1, \ldots, m_t, m)$.

c. Show that $((m_1, \ldots, m_t) : m) = (n_1, \ldots, n_t)$.

d. Show by induction on t that $T(m_1, \ldots, m_t)$ is a resolution of $S/(m_1, \ldots, m_t)$.

One can use the same formulas to define a complex $T(m_1, \ldots, m_t)$ for more general elements m_i of more general rings; essentially all one needs is that the intersection of any subset of the ideals (m_i) is again a principal ideal m_I—this says that the least common multiple operation is well defined. For an interesting example where this complex is exact, see Yuzvinsky [1994].

Conormal Sequence of a Complete Intersection

Exercise 17.12: Let A be an affine ring over a perfect field k. Assume A is equidimensional and locally regular, and let $I \subset A$ be a radical ideal generated by a regular sequence of length c. Let $B = A/I$. Prove using the steps below that I/I^2 is a free B-module and that the map d in the conormal sequence

$$I/I^2 \xrightarrow{d} B \otimes_A \Omega_{A/k} \to \Omega_{B/k} \to 0$$

is injective; thus $\Omega_{B/k}$ has projective dimension ≤ 1 in this case (it is a conjecture of Vasconcelos, recently proved in the graded case by Avramov and Herzog [1994], that this is the only situation in which $\Omega_{B/k}$ has finite projective dimension).

a.* Let A be any ring, I an ideal of A generated by a regular sequence $x_1, \ldots, x_c$. Show that the images in I/I^2 of $x_1, \ldots, x_c$ form a free basis of I/I^2 over A/I.

b. With hypotheses as in part a, suppose in addition that A/I is reduced. Show that the associated primes of I/I^2 are the minimal primes of I.

c. Returning to the situation of the exercise, it is enough by part b to prove that d is an injection locally at each minimal prime P of B. But B_P is a field. Prove the necessary result by comparing the ranks of the free B_P-modules $(I/I^2)_P$, $B_P \otimes_A \Omega_{A/k}$ and $\Omega_{B_P/k}$.

Regular Sequences Are Like Sequences of Variables

Exercise 17.13 (Ideals of monomials in a regular sequence):* Let $y_1, \ldots, y_r$ be a sequence of elements of a ring R such that every subset of $\{y_1, \ldots, y_r\}$ forms a regular sequence in some order. Show that ideals generated by monomials in the y_i obey the same rules for intersections, ideal quotients with each other, and resolutions, as do ideals generated by monomials in the variables of a polynomial ring, as follows (the proof follows Eagon and Hochster [1974]):

a. First consider the "generic case": That is, let $S = \mathbf{Z}[x_1, \ldots, x_r]$, where the x_i are indeterminates. If $J \subset S$ is an ideal generated by monomials in the x_i, show that S/J has a filtration by modules of the form $S/(x_{i_1}, \ldots, x_{i_s})$.

b. Regard R as an S-algebra by the map sending $x_i \mapsto y_i$. Use the Koszul complex to show that $\operatorname{Tor}_i^S(S/(x_{i_1}, \ldots, x_{i_s}), R) = 0$.

c. Use induction on the number of terms in the filtration given in part a to show that $\operatorname{Tor}_i^S(S/J, R) = 0$.

d. If I is any ideal generated by monomials in the y_i, show that $I = JR$ for some monomial ideal J of S. Now show that the Taylor complex may be used to give a free resolution of R/I.

Show that if $I = JR$ and $I' = J'R$ for monomial ideals $J,\ J' \subset S$, then

e. $I \cap I' = (J \cap J')R$.

f. $(I : I') = (J : J')R$.

Blowup Algebra and Normal Cone of a Regular Sequence

Exercise 17.14 (Blowup of a regular sequence):* If $x_1, \ldots, x_r$ is a regular sequence in R, and I is the ideal $I = (x_1, \ldots, x_r)$, show that

a. The natural map Symm $(I) \to \mathcal{B}(I, R) = \oplus_{j=0}^{\infty} I^j$ from the symmetric algebra of I to the blowup algebra of I is an isomorphism.

b. The natural map $R[t_2, \ldots, t_r] \to K(R)$ sending t_i to x_i/x_1 has the "obvious" kernel, generated by the elements $t_i x_1 - x_i$ for $i = 2, \ldots, r$.

Exercise 17.15:* For any ideal I and element $x \in R$, let $\text{in}_I(x)$ be the class of x in I^n/I^{n+1} where n is the greatest integer such that $x \in I^n$ (set $\text{in}_I(x) = 0$ if there is no such integer). Prove the following:

Proposition 17.17. *Suppose that $(R, \mathfrak{m})$ is a local ring and $x_1, \ldots, x_r \in \mathfrak{m}$. If there is an ideal $I \subset \mathfrak{m}$ such that $\text{in}_I(x_1), \ldots, \text{in}_I(x_r)$ is a regular sequence on $\text{gr}_I R$, then $x_1, \ldots, x_r$ is a regular sequence.*

Exercise 17.16 (Normal cone of a regular sequence): The following characterization of regular sequences is due to David Rees [1957]. The reader may compare the result with Exercise 3.16 to see a measure of the difference between regular sequences and systems of parameters.

a.* Let $I = (x_1, \ldots, x_r) \subset R$. If $x_1, \ldots, x_r$ is a regular sequence, show that the natural map $R/I[y_1, \ldots, y_r] \to \text{gr}_I R$ sending y_i to the class of x_i in I/I^2 is an isomorphism. (This condition may be restated by saying that if $F(X_1, \ldots, X_r) \in R[X_1, \ldots, X_r]$ is a homogeneous form of degree d such that $F(x_1, \ldots, x_r) \in I^{d+1}$, then all the coefficients of F are in I. In particular, I^n/I^{n+1} is a free R/I-module for all n.

b.* If R is local and $x_1, \ldots, x_r$ are elements of the maximal ideal, show that the converse is true as well. (Under some circumstances, it suffices to suppose that I/I^2 is free; see Vasconcelos' theorem, Exercise 20.23.)

c. Suppose that R is local and $x_1, \ldots, x_r$ are elements of the maximal ideal. Show that $x_1, \ldots, x_r$ is a regular sequence iff the following variant of the condition in parts a and b is satisfied: Each homogeneous

form $F(X_1, \ldots, X_r) \in R[X_1, \ldots, X_r]$ such that $F(x_1, \ldots, x_r) = 0$ has its coefficients in $(x_1, \ldots, x_r)$.

Exercise 17.17:

a. If M is an R-module and $x \in M$, show that cycles of $K(x)$ (that is the kernel of d_x) is a $\wedge M$-submodule of $\wedge M$ and similarly for the boundaries (the image of d_x). Thus there is a natural $\wedge M$-module structure on the homology $H^*(K(x)) := \oplus_i H^i(K(x))$ of $K(x)$.

b. If N is an R-module and $\varphi : N \to R$ is a homomorphism, use the fact that δ_φ is an derivation of $\wedge N$ to show that the cycles of $K'(\varphi)$ form a subalgebra of $\wedge N$, and the boundaries are an ideal in this algebra. Thus there is a natural algebra structure on the homology $H_*(K'(\varphi)) := \oplus_i H_i(K'(\varphi))$.

c*. Let $M \cong R^2$ have basis m_1, m_2 and let $x = x_1 m_1 + x_2 m_2 \in M$. Let $N \cong R^2$ have basis n_1, n_2 and let $\varphi : N \to R$ be given by $\varphi(n_1) = x_1, \varphi(n_2) = x_2$. Compute the $\wedge M$-module structure on $H^*(K(x))$ and the algebra structure on $H_*(K'(\varphi))$ in case $R = k[x_1, x_2]/(x_1^3, x_2^3)$ and in case $R = k[x_1, x_2]/(x_1, x_2)^3$.

Geometric Contexts of the Koszul Complex

The exercises 17.18–17.20 are for those with some acquaintance with line bundles and sheaves on an algebraic variety (the reader with still more experience may substitute the word "scheme" for "variety" everywhere).

Exercise 17.18 (Castelnuovo's base point free pencil trick): Let X be an algebraic variety over a field k, let $\mathcal{L}$ be an invertible sheaf (line bundle) on X, and let V be a pencil of sections of $\mathcal{L}$: That is, V is a two-dimensional vector space over k of global sections of the line bundle $\mathcal{L}$.

a.* Show that there is a complex

$$0 \to \mathcal{L}^{-1} \to V \otimes \mathcal{O}_X \to \mathcal{L} \to 0$$

that in terms of a local trivialization $\mathcal{L}|_U \cong \mathcal{O}_{X|U}$ is the Koszul complex of the map of modules $V \otimes \mathcal{O}_{X|U} \to \mathcal{O}_{X|U}$. Show that if the pencil is **base-point free** in the sense that the sections in V generate $\mathcal{L}$ locally everywhere, then this complex is exact.

b. Let F be any sheaf on X, and consider the multiplication map $\mu : V \otimes H^0(X; F) \to H^0(X; \mathcal{L} \otimes F)$. Suppose that V is base-point free. Show that if $H^1(\mathcal{L}^{-1} \otimes F) = 0$, then μ is surjective by tensoring F with the exact sequence of part a.

c.* Here is a typical application of part b: Suppose X is a reduced irreducible curve and that $\mathcal{L}$ is generated by global sections and nonspecial ($H^1(X; \mathcal{L}) = 0$). Show that the k-algebra $\oplus_{n \geq 0} H^0(X; \mathcal{L}^{\otimes n})$ is generated in degrees 1 and 2.

Exercise 17.19 (Koszul cohomology): Let X be a variety over a field k, let L be a line bundle on X, let V be any finite-dimensional vector space of sections of L, and let F be any quasicoherent sheaf (that is, sheaf of $\mathcal{O}_X$-modules) on X. Define a complex of sheaves on X:

$$K : \cdots \to \wedge^{i+1} V \otimes \mathcal{L}^{\otimes(j-1)} \otimes F \to \wedge^i V \otimes \mathcal{L}^{\otimes j} \otimes F \to \cdots$$

such that for each affine open set U with affine ring R such that $\mathcal{L}$ is trivial on U we have $K|_U \cong K'(x_1, \dots, x_r) \otimes M$, where $x_1, \dots, x_r$ form a basis of V, regarded as elements of $\mathcal{O}_X(U)$ by means of an identification $\mathcal{L}|U = \mathcal{O}_{X|U}$ and $M = F(U)$, regarded as an $\mathcal{O}_X(U)$-module. Show that if the sections in V do not all vanish simultaneously, then the complex K is exact. If we apply the functor H^0, however, we get a complex

$$H^0(K) : \ \cdots \to \wedge^i V \otimes H^0(\mathcal{L}^{\otimes j} \otimes F) \to \cdots$$

that is generally not exact. Its homology is called the **Koszul cohomology** of $V, \mathcal{L}, F$. This notion has been developed and extensively exploited by Mark Green; see [1984a and b and 1987].

Exercise 17.20 (Zero locus of a section of a vector bundle): Let E be a vector bundle (= locally free sheaf) of rank r on a variety X, and let σ be a section of E.

a. Regarding σ as a map $F := E^* \to \mathcal{O}_X$, form a Koszul complex

$$K : 0 \to \wedge^r F \to \cdots \to \wedge^i F \to \wedge^{i-1} F \to \cdots \xrightarrow{\sigma} F \to \mathcal{O}_X.$$

b. Now suppose that the zero locus Z of σ has codimension r, the "expected" value. Suppose that every local ring $\mathcal{O}_{X,x}$ has the property that any sequence of elements $x_1, \dots, x_r$ generating a codimension-r ideal is a regular sequence (X is "locally Cohen-Macaulay"; see Chapter 18). Show that K is a resolution of the sheaf $\mathcal{O}_Z$ by vector bundles.

Exercise 17.21 (The tautological Koszul complex): Let k be a ring and let V be a finitely generated free module over k. Let $S = S(V)$ be the symmetric algebra of V. There is a natural isomorphism $V^* \otimes V \to \operatorname{Hom}(V, V)$ sending $\varphi \otimes v$ to the map $w \mapsto \varphi(w)v$. Let $t \in V^* \otimes V$ be the element corresponding to $1 \in \operatorname{Hom}(V, V)$. In terms of a basis $\{v_i\}$ of V and a dual basis $\{\varphi_i\}$ of V^*, show that $t = \sum \varphi_i \otimes v_i$.

a. Regarding $t \in V^* \otimes V = \wedge^1(V^*) \otimes S_1(V)$ as an element of the algebra $\wedge(V^*) \otimes S(V)$, show that $t^2 = 0$. Use this to derive the existence of a complex of free $S(V)$-modules

$$K : S(V) \to V^* \otimes S(V) \to \cdots \to \wedge^d(V^*) \otimes S(V) \to$$

where the maps are all given by multiplication by t, called the **tautological Koszul complex**. Show that this is naturally isomorphic

to the Koszul complex $K(t)$, where we regard t as an element of the $S(V)$-module $V^* \otimes S(V)$.

b. Show that the dual of the complex defined in part a is a free resolution of k over $S(V)$. Thus the homology of $k \otimes K$ is isomorphic to $\text{Ext}^*_{S(V)}(k, k)$. Show that the differential of the complex $k \otimes K$ is 0, so the homology is $k \otimes K = \wedge V$. Use Exercise A3.28 to show that the isomorphism is an isomorphism of algebras, where $\text{Ext}^*_{S(V)}(k, k)$ is an algebra with the Yoneda product.

c. Above we decomposed the algebra $\wedge(V^*) \otimes S(V)$ in a certain way to obtain K. Decompose it in a different way to define a complex of free $\wedge V$-modules

$$L : \wedge(V^*) \to \wedge V^* \otimes V \to \cdots \to \wedge(V^*) \otimes S_d(V) \to$$

and show, using the same ideas as for part b, that $\text{Ext}^*_{\wedge(V^*)}(k, k) \cong S(V)$. L is a sort of Koszul complex for $\wedge(V^*)$.

Exercise 17.22 (Priddy's generalized Koszul complex): There is a generalization of the tautological Koszul complex due to Stuart Priddy [1970] that plays a significant role in both commutative and noncommutative algebra. We present an interesting special case. Let k be a field and let $T = T(V)$ be the tensor algebra over k on a finite-dimensional vector space V (that is, T is the free algebra generated by a basis of V). A **quadratic algebra** is a (not necessarily commutative) k-algebra of the form T/I, where I is generated as a two-sided ideal by a vector space of **quadratic relations** $Q \subset V \otimes V = T_2$—that is, $I = TQT$.

a. Show that the symmetric algebra and exterior algebra are quadratic algebras:

$$S(V) = T/I \quad I \text{ is generated by } \{v \otimes w - w \otimes v | v, w \in V\}$$
$$\wedge V = T/I \quad I \text{ is generated by } \{v \otimes w + w \otimes v, v \otimes v | v, w \in V\}.$$

In particular the commutative quadratic algebras are those of the form $S(V)/J$, where J is an ideal generated by quadratic forms.

b. We may identify $V^* \otimes V^*$ with $(V \otimes V)^*$ by the rule $\varphi \otimes \psi(v \otimes w) = \varphi(v)\psi(w)$ for all φ, $\psi \in V^*$ and v, $w \in V$. For any subspace $Q \subset V \otimes V$, we define the **perpendicular subspace** to be $Q^\perp = \{\alpha \in V^* \otimes V^* | \alpha(q) = 0 \text{ for all } q \in Q\}$. If $I \subset T$ is generated by $Q \subset T_2$, let $I^\perp \subset T(V^*)$ be the ideal generated by $Q^\perp \subset T_2(V^*)$; and if $A = T/I$, set $A^\perp = T(V^*)/I^\perp$. Show that

$$A^{\perp\perp} = A$$
$$S(V)^\perp = \wedge(V^*).$$

c. Let $t \in V \otimes V^*$ be as in Exercise 17.21. Regarding t as an element of degree (1,1) of the algebra $A \otimes_k A^\perp$, show that $t^2 = 0$ by considering $t^2 \in A_2 \otimes A_2^\perp = \operatorname{Hom}(A_2^{\perp *}, A_2)$ as a homomorphic image of the element $1 \otimes 1 \in \operatorname{Hom}(V \otimes V, V \otimes V)$. (In fact, $T(V^*)Q^\perp T(V^*)$ is the smallest ideal we could factor out and have this be true!)

d. Deduce the existence of a complex of free A-modules

$$P(A) : A \to A \otimes_k V^* \to A \otimes_k A_2^\perp \to \cdots \to A \otimes_k A_d^\perp \to \cdots,$$

the **Priddy complex**, where the differential is multiplication on the right by t.

e. Show that in case $A = S(V)$, the Priddy complex is the tautological Koszul complex of Exercise 17.21.

f. Show by using Exercise A3.28 that there is a map of algebras $A^\perp = k \otimes_A P(A) \to \operatorname{Ext}_A^*(k,k)$, where $\operatorname{Ext}_A^*(k,k)$ is an algebra by means of the Yoneda product.

It is known that this map is always an injection, and that the image may be characterized either as the subalgebra of $\operatorname{Ext}_A^*(k,k)$ generated by $\operatorname{Ext}_A^1(k,k)$ or as the subalgebra $\oplus^d \operatorname{Ext}_A^d(k,k)_d$ consisting of elements of bidegree (d,d). See Löfwal [1986] for details. The algebra A is called a **Koszul algebra** if $A^\perp = \operatorname{Ext}_A^*(k,k)$. Many algebras of interest in algebraic geometry have this property—roughly speaking it is true of the homogeneous coordinate ring of any variety embedded in projective space by a "sufficiently" ample line bundle—for example, by a sufficiently high multiple of any line bundle. It is also true of the homogeneous coordinate ring of many homogeneous spaces such as the Grassmannian, flag manifold, etc. See Kempf [1990] or Eisenbud, Reeves, and Totaro [1994] and the references there for more information. Among noncommutative algebras, the most interesting examples are perhaps the coordinate rings of quantum groups; see, for example, Manin [1988].

g. We may alternately define the dual of the Priddy complex directly: This will give us a resolution of k as an A-module iff A is Koszul iff K has a resolution in which the dth free module has all its generators in degree d. With $Q \subset V \otimes V$ the space of quadratic relations of A, set

$$\begin{aligned} F_d &= \ker(T_d V \to \sum_{i+j=d-2} T_i V \otimes A_2 \otimes T_j V) \\ &= \cap_{i+j=d-2} (T_i V \otimes Q \otimes T_j V) \subset T_d V. \end{aligned}$$

Show that $F_d = V \otimes F_{d-1} \cap F_{d-1} \otimes V$. This gives a natural map $F_d \to V \otimes F_{d-1}$, which extends to a map of free A-modules $A \otimes F_d \to$

$A \otimes F_{d-1}$. Show that the composite map $F_d \to V \otimes F_{d-1} \to A_2 \otimes F_{d-2}$ is 0, and deduce the existence of the complex

$$P^*(A): \quad \cdots \to A \otimes F_d \to A \otimes F_{d-1} \to \cdots \to A \otimes Q \to A \otimes V \to A.$$

Show that this is the dual of the Priddy complex $P(A)$.

18
Depth, Codimension, and Cohen-Macaulay Rings

In this chapter all the rings are assumed to be Noetherian.

In this chapter we shall use the tools forged in Chapter 17 to explore the basic facts about **Cohen-Macaulay rings**, which are rings R in which $\operatorname{depth}(I, R) = \operatorname{codim} I$ for every ideal I (it is enough to assume this when I is a maximal ideal). These rings are important because they provide a natural context, broad enough to include the rings associated to many interesting classes of singular varieties and schemes, to which many results about regular rings can be generalized. The geometric meaning of the Cohen-Macaulay property is somewhat obscure but has a good expression in terms of maps to regular varieties, as we shall see in Theorem 18.16 and Corollary 18.17.

Perhaps the most significant of the results of this section is the unmixedness theorem (Corollary 18.14), which explains, for example, why hyperplane sections of smooth varieties do not have embedded components. This is the reason that we can treat divisors as codimension 1 subvarieties. It is thus a pillar of algebraic geometry. (For those who know about schemes, the unmixedness theorem says the same thing for any Cartier divisor on a locally Cohen-Macaulay scheme, and still more generally for any scheme satisfying the property S2.)

18.1 Depth

Recall from the previous chapter that if I is an ideal of a ring R, and M is a finitely generated R-module such that $IM \neq M$, then the depth of I on M,

written depth(I, M), is the length of a (indeed any) maximal M-sequence in I. When $M = R$, we shall simply speak of the depth of I. Theorem 17.4 characterizes depth(I, M) in terms of the vanishing of the homology of the Koszul complex.

As usual, we shall frequently want to localize, so a remark on the behavior of depth under localization is in order.

Lemma 18.1. *If R is a ring, and P is a prime ideal in the support of a finitely generated R-module M, then any M-sequence in P localizes to an M_P-sequence. Thus for any ideal $I \subset P$ we have* depth$(I, M) \leq$ depth(I_P, M_P)*, the latter taken in the ring R_P. In general, the inequality may be strict, but for any ideal I there exist maximal ideals P in the support of M such that* depth$(I, M) =$ depth(I_P, M_P)*. In particular, if P is a maximal ideal, then* depth$(P, M) =$ depth(P_P, M_P).

Proof. Nakayama's lemma guarantees that $I_P M_P \neq M_P$, the only tricky part of the first statement. The depth really can increase on localization, since for example the localization map $M \to M_P$ might kill some of the elements killed by elements of I, so that these elements of I might become nonzerodivisors.

For the second statement write $I = (x_1, \ldots, x_n)$, and set $r =$ depth(I, M). By Theorems 17.4 and 17.6, $H^r(M \otimes K(x_1, \ldots, x_n)) \neq 0$ and the primes P containing I such that depth$(I_P, M_P) =$ depth(I, M) are exactly the primes in the support of $H^r(M \otimes K(x_1, \ldots, x_n))$. In particular, there are some maximal ideals P with this property. The last statement of the lemma follows at once from the second.

For an alternate approach to the second statement, which does not require the Koszul complex, choose a maximal M-sequence $x_1, \ldots, x_r$ in I. Because I consists of zerodivisors on $M/(x_1, \ldots, x_r)M$, I is contained in the union of the associated primes of $M/(x_1, \ldots, x_r)M$, and since the set of associated primes is finite, prime avoidance (Lemma 3.3) shows that I is contained in one of them. Localization at this prime, or at any prime containing it, will preserve the depth of I on M.

We have already remarked that depth I is a measure of the size of I, as is codim I. In this section we shall explore the relation between these two notions. First we show that there is always an inequality. It is technically useful to work with the depth of I on a module M. We write ann(M) for the annihilator of M in R. If $x \in R$ then the action of x on M depends only on the residue class of x modulo ann(M), so the depth of I on M is the same as that of $I +$ ann(M) on M. For this reason we can restrict attention to ideals containing ann(M).

Proposition 18.2. *Let R be a ring and let M be a finitely generated R-module. If I is an ideal of R containing* ann(M)*, then* depth(I, M) *is $\leq$ the length of any maximal chain of prime ideals descending from a prime*

containing I to an associated prime of M. In particular, the depth of I (on R) is $\leq \operatorname{codim} I$.

In the proof we use a result that plays a role in the theory of depth something like the one played by the principal ideal theorem in the theory of codimension.

Lemma 18.3. *If R, P is a local ring, M is a finitely generated R-module, I is an ideal of R, and $y \in P$, then*

$$\operatorname{depth}((I, y), M) \leq \operatorname{depth}(I, M) + 1.$$

Proof. Let $x_1, \ldots, x_n$ be a set of generators for I, and set $r = \operatorname{depth}(I + (y), M)$. By Theorem 17.4, $H^i(M \otimes K(x_1, \ldots, x_n, y)) = 0$ for $i < r$. By Corollary 17.11 and Nakayama's lemma, $H^i(M \otimes K(x_1, \ldots, x_n)) = 0$ for $i < r - 1$. By Theorem 17.4, $\operatorname{depth} I \geq r - 1$.

Proof of Proposition 18.2. Though the second statement follows from the first, its proof is so easy as to be worth giving separately: Let $x_1, \ldots, x_n$ be a maximal R-sequence in I. Since x_1 is a nonzerodivisor, it is not contained in any minimal prime of R, so the codim $I/(x_1)$ (as an ideal in $R/(x_1)$) $< \operatorname{codim} I$. But the depth of $I/(x_1)$ as an ideal in $R/(x_1)$ is $n-1$, so by induction $n - 1 \leq \operatorname{codim} I/(x_1) < \operatorname{codim} I$, and we are done.

For the main result, let $Q \supset Q_1 \supset \cdots \supset Q_l$ be any maximal chain of primes descending from a prime Q containing I to a prime Q_l associated to M. We do induction on l. The case $l = 0$, where Q is an associated prime, is immediate from the definitions.

Now suppose $l \geq 1$. Enlarging I, we may as well assume that $I = Q$. If we localize at Q, any regular sequence in Q remains a regular sequence, so the depth can only increase and we may suppose that R is local and that Q is its maximal ideal. Let $x \in Q$ be an element outside Q_1. Since Q is the only prime minimal over $Q_1 + (x)$, Corollary 2.12 shows that Q is nilpotent mod $Q_1 + (x)$. Thus by Corollary 17.8 $\operatorname{depth} Q = \operatorname{depth}(Q_1 + (x))$.

By Lemma 18.3, $\operatorname{depth}(Q_1 + (x)) \leq \operatorname{depth}(Q_1) + 1$. From the inductive hypothesis we get $\operatorname{depth}(Q_1, M) \leq l - 1$. Putting these inequalities together we get $\operatorname{depth} Q \leq \operatorname{depth}(Q_1, M) + 1 \leq l$ as required. □

18.1.1 Depth and the Vanishing of Ext

There is another characterization of depth that generalizes, in a certain sense, the characterization by the homology of the Koszul complex of Theorem 17.4. We shall apply it in Theorem 18.12 and again in Chapter 20. The following result contains our first use of the functor Ext, the analogue of Tor in which the functor $\operatorname{Tor}_0^R(M, N) = M \otimes_R N$ is replaced by the functor $\operatorname{Ext}_R^0(M, N) = \operatorname{Hom}_R(M, N)$. The reader may wish to review the Introduction to Tor section in Chapter 6 and compare it with the material on Ext in Appendix 3 at this point.

Proposition 18.4. *Let R be a ring and let M and N be finitely generated R-modules. If* $\operatorname{ann} M + \operatorname{ann} N = R$ *then* $\operatorname{Ext}^r_R(M, N) = 0$ *for every r. Otherwise,* $\operatorname{depth}(\operatorname{ann}(M), N)$ *is the smallest number r such that* $\operatorname{Ext}^r_R(M, N) \neq 0$.

Proof. Since Ext is an R-linear functor in each variable, $\operatorname{Ext}^*_R(M, N)$ is annihilated by each element of $\operatorname{ann} M$ and $\operatorname{ann} N$ (see Appendix 3), so the first statement is clear.

First we show that $\operatorname{ann} M + \operatorname{ann} N = R$ iff $\operatorname{ann}(M)N = N$. Suppose that $\operatorname{ann}(M)N = N$: By Corollary 4.7, there is an element $r \in \operatorname{ann}(M)$ such that $(1 - r)N = 0$. Thus $1 \in \operatorname{ann} M + \operatorname{ann} N$. Conversely, if we can write $1 = r + s$ with $r \in \operatorname{ann} M$ and $s \in \operatorname{ann} N$, then $rN = (r + s)N = N$.

Now suppose that $\operatorname{ann} M + \operatorname{ann} N \neq R$. Since then $\operatorname{ann}(M)N \neq N$, the number $d = \operatorname{depth}(\operatorname{ann}(M), N)$ is $< \infty$, and we do induction on d. If $d = 0$ we must show that $\operatorname{Hom}(M, N) \neq 0$. Since $\operatorname{depth}(\operatorname{ann} M, N) = 0$, there is an associated prime P of N that contains $\operatorname{ann} M$. Since M is finitely generated, the formation of Hom commutes with localization (Proposition 2.10), so it is enough to prove the result after localizing at P. After localizing we are in the situation where R is local, and N contains a copy of the residue class field R/P. Since $M \neq 0$, Nakayama's lemma shows that $M/PM \neq 0$, so M/PM is a nonzero direct sum of copies of R/P. Thus there is a nonzero map $M \to R/P \subset N$.

Next suppose that $d \geq 1$, and let $x \in \operatorname{ann} M$ be a nonzerodivisor on N. We have $\operatorname{ann}(M)N/xN \neq N/xN$, and $\operatorname{depth}(\operatorname{ann}(M), N/xN) = d - 1$. By induction $\operatorname{Ext}^i_R(M, N/xN) \neq 0$ for $i = d - 1$, but for no smaller i.

We apply the long exact sequence in $\operatorname{Ext}_R(M, -)$ to the short exact sequence

$$0 \to N \xrightarrow{x} N \to N/xN \to 0.$$

Since x kills M, it kills each $\operatorname{Ext}^j_R(M, N)$. Thus $\operatorname{Hom}(M, N) = 0$, and we obtain short exact sequences

$$0 \to \operatorname{Ext}^{j-1}_R(M, N) \to \operatorname{Ext}^{j-1}_R(M, N/xN) \to \operatorname{Ext}^j_R(M, N) \to 0$$

for every $j \geq 1$. By induction on i it follows that $\operatorname{Ext}^i_R(M, N) = 0$ for $i < d$, while $\operatorname{Ext}^d_R(M, N) \neq 0$, as required.

The connection with Theorem 17.4 is as follows: If $x_1, \ldots, x_n$ is a regular sequence in R, then the Koszul complex $K(x_1, \ldots, x_n)$ is a free resolution of $R/(x_1, \ldots, x_n)$ by Corollary 17.5. Thus the homology of the complex $\operatorname{Hom}(K(x_1, \ldots, x_n), M)$ is $\operatorname{Ext}_R(N, M)$ where $N = R/(x_1, \ldots, x_n)$. Since the Koszul complex is isomorphic to its own dual, $\operatorname{Hom}(K(x_1, \ldots, x_n), M) \cong M \otimes_R K(x_1, \ldots, x_n)$ as complexes, so Theorem 17.4 coincides with Proposition 18.4 in this case.

Recall that $pd_R(M)$ is the minimum length of a free resolution of M. Since $\operatorname{Ext}(M, N)$ is computed from a free resolution of M, we see that if $\operatorname{Ext}^i_R(M, N) \neq 0$ for any module N, then $pd_R M \geq i$. In particular, we get

Corollary 18.5. *For any nonzero module M, $pd_R M \geq \operatorname{depth} \operatorname{ann} M$.*

Proof. Take $N = R$ in Proposition 18.4.

See Exercise 19.9 for the case of equality, and Exercise 19.13 for a sharper version.

In case (R, P) is local, Proposition 18.4 shows that the depth of any module N (that is, $\operatorname{depth}(P, N)$) may be computed from the vanishing behavior of $\operatorname{Ext}^*_R(R/P, N)$. For every short exact sequence of modules

$$0 \to N' \to N \to N'' \to 0$$

we get a long exact sequence in Ext, so we get inequalities on the depths of N, N', and N''. (We could have done the same thing using the homology of the Koszul complex.) We record two of them for use in Chapter 20.

Corollary 18.6. *With notation as above, if N, N' and N'' are nonzero, then*

a. $\operatorname{depth} N'' \geq \min(\operatorname{depth} N, \operatorname{depth} N' - 1)$.

b. $\operatorname{depth} N' \geq \min(\operatorname{depth} N, \operatorname{depth} N'' + 1)$.

18.2 Cohen-Macaulay Rings

In Corollary 10.15 we showed that if (R, P) is a regular local ring (that is, if the number of generators of P is equal to the dimension of R), then any minimal set of generators for P is a regular sequence. Thus the inequality $\operatorname{depth} P \leq \operatorname{codim} P$ of Proposition 18.2 becomes an equality in this case. It turns out that equality holds in many nonregular rings as well. The following result will help us exploit this equality when it occurs.

Theorem 18.7. *Let R be a ring such that* $\operatorname{depth} P = \operatorname{codim} P$ *for every maximal ideal P of R. If $I \subset R$ is a proper ideal, then* $\operatorname{depth} I = \operatorname{codim} I$.

For a generalization to the case of modules, see Exercise 18.4 and the results around Proposition 21.9.

Proof. By Proposition 18.2 we have $\operatorname{depth}(I) \leq \operatorname{codim}(I)$, and we must prove the other inequality.

By Lemma 18.1 we may localize at some maximal ideal $P \supset I$ without disturbing the depth of I or the depth of P, so we may assume that (R, P) is local with $I \subset P$. If I is P-primary, then $\operatorname{codim} I = \operatorname{codim} P$. By Corollary 17.8, $\operatorname{depth} I = \operatorname{depth} P$, so the theorem is true for I. Thus we may assume that I is not P-primary. By Noetherian induction, we may assume that the theorem holds for all ideals strictly larger than I.

Since P is not a minimal prime of I, we may by prime avoidance (Lemma 3.3) find an element $x \in P$ not in any minimal prime of I. By the induction we have $\operatorname{depth}(I + (x)) = \operatorname{codim}(I + (x)) = \operatorname{codim}(I) + 1$. But by Lemma 18.3 $\operatorname{depth}(I+(x)) \leq \operatorname{depth}(I)+1$, so $\operatorname{depth}(I) \geq \operatorname{codim}(I)$, as required.

Theorem 18.7 is so useful that its hypothesis has become one of the central definitions in commutative algebra.

Definition. *A ring such that* $\operatorname{depth} P = \operatorname{codim} P$ *for every maximal ideal* P *of* R *is called a* **Cohen-Macaulay** *ring.*

The Cohen-Macaulay property is local in a strong sense.

Proposition 18.8. *R is Cohen-Macaulay iff R_P is Cohen-Macaulay for every maximal ideal P of R, and then R_Q is Cohen-Macaulay for every prime Q of R. A local ring is Cohen-Macaulay iff its completion is Cohen-Macaulay.*

Proof. If R is Cohen-Macaulay, and Q is a prime ideal, then $\operatorname{codim} Q_Q = \operatorname{codim} Q = \operatorname{depth}(Q, R) \leq \operatorname{depth} Q_Q \leq \operatorname{codim} Q_Q$ by Proposition 18.2, so the inequality is an equality and R_Q is Cohen-Macaulay. If R_P is Cohen-Macaulay for every maximal ideal P, then $\operatorname{depth}(P, R) = \operatorname{depth}(P_P, R_P)$ by Lemma 18.1. As $\operatorname{codim} P = \operatorname{codim} P_P$, we see that R is Cohen-Macaulay.

Now suppose that (R, P) is a local ring, and let $(\hat{R}, \hat{P})$ be its completion. We already know that $\operatorname{codim} \hat{P} = \operatorname{codim} P$, so it is enough to show that $\operatorname{depth}(\hat{P}, \hat{R}) = \operatorname{depth}(P, R)$. Let $x_1, \ldots, x_n$ be generators for P. From the construction we see that $\hat{R} \otimes_R K(x_1, \ldots, x_n)$ is the Koszul complex $\hat{K}$ of $x_1, \ldots, x_n$ as elements of $\hat{R}$. By Theorem 7.2b, $\hat{R}$ is flat over R so we have $H^*(\hat{K}) = \hat{R} \otimes H^*(K(x_1, \ldots, x_n))$. By Theorem 7.2a and Nakayama's lemma, any finitely generated nonzero R-module remains nonzero on tensoring with $\hat{R}$, so $\operatorname{depth}(\hat{P}, \hat{R}) = \operatorname{depth}(P, R)$ by Theorem 17.4.

The Cohen-Macaulay property passes to polynomial rings:

Proposition 18.9. *A ring R is Cohen-Macaulay iff the polynomial ring $R[x]$ is Cohen-Macaulay.*

Proof. If $R[x]$ is Cohen-Macaulay, then since x is a nonzerodivisor, $R[x]/(x) = R$ is Cohen-Macaulay.

For the converse, it suffices by Proposition 18.8 to prove that each localization of $R[x]$ at a maximal ideal is Cohen-Macaulay. Let P be a maximal ideal of $R[x]$, and let $Q = P \cap R$. Since the complement of Q in R is contained in the complement of P in $R[x]$ we have

$$R[x]_P = R_Q[x]_P,$$

so we may assume that R is local with maximal ideal Q. The ring $R[x]/QR[x] = (R/Q)[x]$ is a principal ideal domain, so modulo Q the ideal P is generated by a monic polynomial $f(x)$: That is, $P = (Q, f(x))$. If $x_1, \ldots, x_n$ is an R-sequence in Q, then it is also an $R[x]$-sequence since $R[x]$ is a free R-module. Further, the monic polynomial $f(x)$ is a nonzero-divisor modulo any ideal of R, so $x_1, \ldots, x_n, f(x)$ is an $R[x]$-sequence. Thus $\operatorname{depth} P \geq 1 + \operatorname{depth} Q$.

On the other hand, $\operatorname{codim} P \leq 1 + \operatorname{codim} Q$ by the principal ideal theorem. Since R is Cohen-Macaulay we have $\operatorname{codim} Q = \operatorname{depth} Q$, and we obtain $\operatorname{codim} P \leq \operatorname{depth} P$. The opposite inequality is immediate from Proposition 18.2, so $R[x]_P$ is Cohen-Macaulay as required.

Now we turn to some of the desirable properties of Cohen-Macaulay rings. The following corollary is a substantial generalization of part of Theorem A from Chapter 8. Recall that a ring R is **catenary**, or **has the saturated chain condition**, if given any primes $P \subset Q$ of R, the maximal chains of primes between P and Q all have the same length. R is **universally catenary** if every finitely generated R-algebra is catenary. It follows at once that a homomorphic image of a universally catenary ring is universally catenary.

Corollary 18.10. *Cohen-Macaulay rings are universally catenary. Moreover, in a local Cohen-Macaulay ring, any two maximal chains of primes have equal length, and every associated prime of R is minimal.*

Proof. Since the polynomial ring over a Cohen-Macaulay ring is again Cohen-Macaulay by Proposition 18.9, it suffices to show that a homomorphic image S of a Cohen-Macaulay ring R is catenary. Any two maximal chains between a given pair of primes in S pull back to two maximal chains between two primes $Q \subset P$ in R. By Proposition 18.8 we may localize and suppose that R is local with maximal ideal P. The two chains may be extended to maximal chains in R by adding the same chain of primes descending from Q to each. Thus the first statement of the corollary will follow from the second statement.

Let (R, P) be a local Cohen-Macaulay ring. For the second statement of the corollary it is sufficient to show that all maximal chains of primes from P to an associated prime of R have the same length, namely dim R. By Proposition 18.2, the length of any such chain $\geq \operatorname{depth} P$. But $\operatorname{depth} P = \dim P$, the maximal length of such a chain, by hypothesis.

There is a corresponding statement for modules; see Exercise 18.5.

A little more is true in the direction of Corollary 18.10. To formulate it, define a ring to be **equidimensional** if all its maximal ideals have the same codimension and all its minimal primes have the same dimension. For a Cohen-Macaulay ring, the latter condition follows from the former by Corollary 18.10.

If R_1 and R_2 are Cohen-Macaulay rings, then by Exercise 18.6, the direct product $R_1 \times R_2$ is also Cohen-Macaulay. It follows that a Cohen-Macaulay ring need not be equidimensional. However, from Corollary 18.10 we immediately derive:

Corollary 18.11. *Any local Cohen-Macaulay ring is equidimensional.*

Geometrically, Corollary 18.11 says that if a variety X is locally Cohen-Macaulay at a point p (in the sense that the local ring $\mathcal{O}_{X,p}$ is Cohen-Macaulay), then p cannot lie on two components of different dimensions. Thus, for example, the affine ring of the variety shown in Figure 18.1, or even its localization at the singular point, is not Cohen-Macaulay.

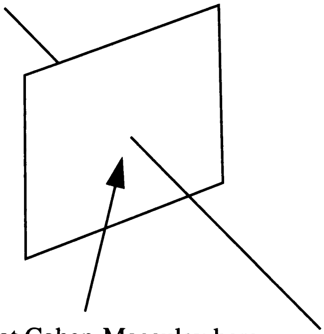


FIGURE 18.1.

The property established in Corollary 18.11 is essentially topological (either in the classical topology, if we work with varieties over **C**, or in the Zariski topology). Unfortunately, no complete characterization of Cohen-Macaulayness in topological terms is known, but there is one more restriction of the topology of a variety near a Cohen-Macaulay point: A result of Hartshorne [1962] says that at a Cohen-Macaulay point, a variety (or scheme) must be locally (analytically) "connected in codimension 1" in the sense that removing a subvariety of codimension 2 or more cannot disconnect it (or even disconnect the spectrum of its completion). Algebraically, we may state the result (in a somewhat strengthened form) as follows:

Theorem 18.12 (Hartshorne's Connectedness Theorem). *Let R be a local ring and let I and J be proper ideals of R whose radicals are incomparable. If $I \cap J$ is nilpotent, then* $\operatorname{depth}(I + J) \leq 1$. *In particular, if R is a Cohen-Macaulay ring, or even satisfies Serre's condition* S2, *then* $\operatorname{codim}(I + J) \leq 1$.

Proof. Replacing I and J by I^N and J^N for a suitable integer $N \gg 0$, and using the fact that $IJ \subset I \cap J$ is nilpotent, we may assume $IJ = 0$. Therefore, $(I \cap J)(I + J) = 0$. Thus if $I \cap J \neq 0$, then $I + J$ consists of zerodivisors and has depth 0.

Thus we may assume $I \cap J = 0$. Consider the short exact sequence

$$0 \to R \to R/I \oplus R/J \to R/(I+J) \to 0,$$

where the image of 1 in $R = R/(I \cap J)$ is $(1, -1) \in R/I \oplus R/J$, and the right-hand map sends (a, b) to $a + b$. If $\operatorname{depth}(I + J) > 1$, then by Proposition 18.4, $\operatorname{Ext}^1_R(R/(I+J), R) = 0$, so the sequence splits, and there is an element $(a, b) \in R/I \oplus R/J$ such that $a+b \equiv 1 \bmod (I+J)$ and such that $R(a, b) \cap R(1, -1) = 0$ in $R/I \oplus R/J$.

Since R is local, either a or b —say a —must be a unit. Thus $J(a, b) = (Ja, 0) = (J, 0) \subset R(a, b)$. But $J(1, -1) = (J, 0)$, so we must have $(J, 0) = 0$, that is, $J \subset I$. It follows that the radical of J is contained in the radical of I, contradicting our assumption.

Thus, for example, a variety that looks locally like two surfaces meeting in a point in four-space, as suggested in Figure 18.2, cannot be Cohen-Macaulay.

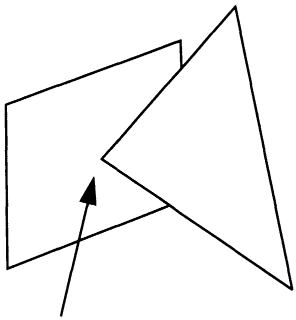


FIGURE 18.2.

Corollary 18.10 and Theorem 18.12 may be used to get information about an ideal I in a ring R as soon as we know that R/I is Cohen-Macaulay. The simplest case in which this hypothesis is satisfied is given by the following result.

Proposition 18.13. *Let R be a Cohen-Macaulay ring. If $I = (x_1, \ldots, x_n)$ is an ideal generated by n elements in a Cohen-Macaulay ring R such that* $\operatorname{codim} I = n$, *the largest possible value, then R/I is a Cohen-Macaulay ring.*

Proof. By Proposition 18.8 we may assume that R is local, with maximal ideal P. Choosing a maximal regular sequence in P that begins with $x_1, \ldots, x_n$, we see that $\operatorname{depth} R/I = \operatorname{depth} R - n$. On the other hand, $\dim R/I \leq \dim R - n$ because for $i = 1, \ldots, n$, the element x_i is not in any of the minimal primes of $(x_1, \ldots, x_{i-1})$.

A consequence, the second statement of the following corollary, was proved by Macaulay in the case R is a polynomial ring, and by Cohen for regular local rings. This is the reason for the name "Cohen-Macaulay."

Corollary 18.14 (Unmixedness Theorem). *Let R be a ring. If $I = (x_1, \ldots, x_n)$ is an ideal generated by n elements such that* $\operatorname{codim} I = n$, *then all minimal primes of I have codimension n. If R is Cohen-Macaulay, then every associated prime of I is minimal over I.*

Proof. Since the codimension of I is the minimum of the codimensions of the minimal primes, we see that these all have codimension $\geq n$. By the principal ideal theorem, they all have codimension $\leq n$.

If now R is Cohen-Macaulay, then R/I is Cohen-Macaulay by Proposition 18.13, so all the associated primes of I (that is, the primes of $\operatorname{Ass}(R/I)$) are minimal over I by Corollary 18.10.

A common geometric use of the unmixedness theorem is to verify that a given set of polynomials generates the homogeneous coordinate ring of a given projective variety. Perhaps the oldest example is given in Exercise 18.10.

For those with some geometric background, here is another, more typical application of the same kind: Let k be an algebraically closed field. Consider a curve C of genus 1 and degree 4 in $\mathbf{P}^3(k)$ (an elliptic quartic). We examine the homogeneous ideal I of C. The curve does not lie in a plane, since a plane quartic has genus 3. By Riemann-Roch we have $h^0(C; \mathcal{O}_C(2)) =$ degree $\mathcal{O}_C(2) = 8$, while the space of quadrics in $\mathbf{P}^3$ is 10-dimensional. Thus there exist two linearly independent quadrics Q_1, Q_2 that vanish on C. We claim that they form a regular sequence and generate the homogeneous ideal of C.

First we show that Q_1 must be irreducible. Suppose on the contrary that $Q_1 = LL'$, the product of two linear forms. Since C is irreducible it would lie in one of the planes $L = 0$ or $L' = 0$, and we have already seen that C does not lie in a plane. Next, since the homogeneous coordinate ring $S = k[x_0, x_1, x_2, x_3]$ of $\mathbf{P}^3$ is factorial, the irreducibility of Q_1 implies that (Q_1) is a prime ideal. Thus Q_2 is a nonzerodivisor $\operatorname{mod} Q_1$; that is, Q_1, Q_2 is a regular sequence. Now $S/(Q_1, Q_2)$ has degree 4 by Bezout's theorem. Since (Q_1, Q_2) is contained in I, we must have $(Q_1, Q_2) = I \cap J$, where $\operatorname{codim} J > 2$. By Corollary 18.14 the ideal (Q_1, Q_2) has no primary components of codimension > 2, so $I = (Q_1, Q_2)$.

18.3 Proving Primeness with Serre's Criterion

Given elements $f_1, \ldots, f_n$ in a ring S, it is often of interest to know whether the ideal $I = (f_1, \ldots, f_n)$ is prime. There are some special methods, like Eisenstein's criterion (Exercise 18.11), and there is a general method that seeks to identify S/I with an explicitly known subring of a domain, perhaps by identifying a vector space basis for each, as in the case given in Exercise 1.19. But in general the problem is quite hard, even when S is as simple as a polynomial $S = k[x_1, \ldots, x_r]$ ring over a field k.

Roughly speaking, there are two aspects of primeness of an ideal $I \subset S$. First, geometrically, the variety corresponding to I must be irreducible (not the union of two smaller varieties). Second, arithmetically, I must be a radical ideal; that is, I must have no embedded components, and must be equal to its radical at the generic points of its isolated components. In many common circumstances the geometric condition is relatively easy to check—often one starts with an irreducible locus and has enough equations to cut it out set-theoretically—but the arithmetic condition is obscure. However, if $R := S/I$ is Cohen-Macaulay, then the fact that R has "many" nonzerodivisors helps with the arithmetic condition. For example, by the unmixedness theorem, the Cohen-Macaulay condition guarantees that there will be no embedded components. The last point is to know that there will be no isolated nonradical components. We shall give a method for checking this based on the Jacobian criterion.

Sometimes one does not know the locus of points where the ideal I vanishes, and therefore one does not know in advance that it is irreducible. In such cases, at least when the variety described by I is actually normal, Serre's criterion allows one to deduce the primeness of I from the Jacobian criterion directly. The point here is that if X is a connected variety (for example, one given by homogeneous equations, so that X is a cone), then the points where two components of X meet are singular points of X and are in the singular locus. Hartshorne's connectedness theorem shows that if R is Cohen-Macaulay then the components must meet in codimension 1 in X if they meet at all.

Before stating the result, one further remark is in order: Since the methods we shall employ are local, we cannot hope to do more than prove that $R = S/I$ is "locally a domain." By Proposition 2.20 this is equivalent to the statement that R is a direct product of domains. In applications where we wish to show that R is actually a domain, we shall need some further information; for example, we might know the variety X as a point set and know that it is connected; or we might know that R cannot have nontrivial idempotents because it is local (so that any nonzero idempotent a unit) or graded (so that any idempotent has degree 0) with degree-0 part a domain.

Theorem 18.15. *Let $R = k[x_1, \ldots, x_r][U^{-1}]/I$ be a localization of an affine ring over a perfect field k. Suppose that $I = (f_1, \ldots, f_n)$ has codimension c.*

Let $J \subset R$ be the ideal generated by the $c \times c$ minors of the Jacobian matrix $\mathfrak{J} = (\partial f_i/\partial x_j)$, taken modulo I. Suppose R is Cohen-Macaulay.

a. *R is reduced iff J has codimension ≥ 1 in R.*

b. *R is a direct product of domains iff condition a holds and R_P is a domain for every prime P of codimension ≤ 1.*

c. *R is a direct product of normal domains iff J has codimension ≥ 2 in R.*

Proof. Parts a and c will be a direct application of the Jacobian criterion together with Serre's criterion and the related criterion of Exercise 11.10. To see this, we first reinterpret the conditions S1 and S2 that appear in Exercise 11.12 and Serre's criterion.

The condition S1 is that all primes associated to 0 have codimension 0. Now a prime P is associated to 0 in R iff P_P is associated to 0 in R_P iff $\operatorname{depth} R_P = 0$. (Here as usual we write depth R_P for $\operatorname{depth}(P_P, R_P)$.) Thus we may restate S1 as the condition

S1$'$: For every prime P of codimension ≥ 1 we have depth $R_P \geq 1$.

The condition S2 comprises the condition S1 as well as the requirement that if $(x) \subset R$ is an ideal generated by a nonzerodivisor and P is an associated prime of (x), then the codimension of P is 1. We claim that these conditions are the same as the conditions S1$'$ and

S2$'$: For every prime P of codimension ≥ 2 we have depth $R_P \geq 2$.

Indeed, suppose that R satisfies S1$'$ and S2$'$. We have seen that R satisfies S1$'$ Let $x \in R$ be a nonzerodivisor and P a prime associated to (x); we must show that P has codimension ≤ 1. But P_P is then associated to xR_P, so depth $R_P \leq 1$, and P must have codimension ≤ 1 by S2$'$.

Conversely, suppose that R satisfies S2, and that P is a prime of codimension ≥ 2. If $\operatorname{depth} R_P \leq 1$, then since $1 \geq \operatorname{depth}(P_P, R_P) \geq \operatorname{depth}(P, R)$ we would either have $\operatorname{depth} P = 0$ (so that P would have codimension 0 by S1) or $\operatorname{depth} P = 1$. In the latter case, P contains a nonzerodivisor x and consists of zerodivisors modulo x. Thus P is contained in an associated prime Q of (x). By S2, $\operatorname{codim} Q \leq 1$, and it follows that $\operatorname{codim} P$ would be less than 1. Since these possibilities contradict our hypothesis, we have depth $R_P \geq 2$.

Now if R is Cohen-Macaulay, then by Proposition 18.8, $\operatorname{depth} R_P = \dim R_P = \operatorname{codim} P$ for every prime P of R; thus R satisfies S2 (and of course S1).

Returning to the hypotheses of Theorem 18.15, R is equidimensional because it is Cohen-Macaulay, and we may apply the Jacobian criterion, Theorem 16.19, to compute the singular locus. The separability hypothesis is guaranteed since we have assumed that k is perfect. Thus if P is a prime

of R then R_P is nonsingular iff $J \not\subset P$. The condition R0 thus says that J has codimension ≥ 1, while the condition R1 says that J has codimension ≥ 2. Thus parts a and c follow at once from the criteria of Exercise 11.10 and Serre's criterion, Theorem 11.5, respectively.

For part b suppose first that R is a product of domains. Every localization of R is then a domain and R is reduced, so the conditions of part b hold. Conversely, suppose that part a is satisfied (so that R is reduced) and R_P is a domain for each prime P of codimension ≤ 1. We must show that R is locally a domain. All the hypotheses are preserved by localization, so we may as well assume that R is local from the outset. Since R is reduced, it suffices to show that R has only one minimal prime. For this purpose we apply Hartshorne's connectedness theorem. If $Q_1, \ldots, Q_s$ were the minimal primes of R, then taking $I = Q_1$ and $J = Q_2 \cap \cdots \cap Q_s$ in Theorem 18.12, we see that $\operatorname{depth}(I + J) \leq 1$. Since R is Cohen-Macaulay we also have $\operatorname{codim}(I + J) \leq 1$. Thus there is a prime P of codimension 1 containing both I and J, and it follows that P contains two minimal primes. But this contradicts the hypothesis that R_P is a domain. The contradiction proves that there cannot be two minimal primes of R, and shows that R is a domain.

Here is a fairly typical example of how this result is used. For a more elementary case see Exercise 18.12.

Example. As an example, consider again the curves cut out by two quadrics in $\mathbf{P}^3$. This time, let us reverse the process and start with an ideal generated by two quadrics, $I = (Q_1, Q_2) \subset S = k[x_0, x_1, x_2, x_3]$, with k algebraically closed and of characteristic not 2, say. Suppose that the Q_i are given explicitly, for example, as

$$\begin{aligned} Q_1 &= x_0^2 + x_1^2 + x_2^2 + x_3^2 \\ Q_2 &= a_0 x_0^2 + a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 \quad \text{with } a_0, \ldots, a_3 \in k. \end{aligned}$$

For what values of $a = (a_0, \ldots, a_3)$ is I prime of codimension 2? For what values is it the ideal of a nonsingular curve? (It is easy to see geometrically that the curve is then an elliptic quartic, as in this example.)

Since Q_1 is irreducible and S is factorial, (Q_1) is prime, so Q_1, Q_2 is a regular sequence as soon as Q_2 is not a multiple of Q_1—that is, as soon as the a_i are not all equal. If all the a_i are equal, then $(Q_1, Q_2) = (Q_1)$ is certainly not a curve, so we shall henceforward assume that not all the a_i are equal. Thus $R := S/I$ is Cohen-Macaulay by Proposition 18.13, and we may apply Theorem 18.15.

The Jacobian matrix of the two equations Q_1, Q_2 is

$$\mathfrak{J} = \begin{pmatrix} 2x_0 & 2x_1 & 2x_2 & 2x_3 \\ 2a_0x_0 & 2a_1x_1 & 2a_2x_2 & 2a_3x_3. \end{pmatrix}$$

Write $J' \subset S$ for the ideal generated by the 2×2 minors of $\mathfrak{J}$. Since 2 is a unit, the Jacobian ideal J' is generated by the elements $(a_i - a_j)(x_i x_j)$ for $i \neq j$.

Suppose first that two of the a_i are equal, say $a_0 = a_1$. The element $Q_2 - a_0 Q_1 = (a_2 - a_0)x_2^2 + (a_3 - a_0)x_3^2$ factors as $(b_2 x_2 + b_3 x_3)(b_2 x_2 - b_3 x_3)$, where b_i is a square root of $a_i - a_0$. The factors are linear, and thus certainly not contained in I, so in this case I is not a prime, and we may assume that all the a_i are distinct.

With this hypothesis J' is the ideal generated by all the $x_i x_j$ for $i \neq j$. This monomial ideal has codimension 3, so we see at once by Theorem 18.15a that R is reduced. Since J' is monomial, it is easy to compute its minimal primes: They are primes generated by the minimal subsets of variables necessary to generate ideals containing J'. Thus the minimal primes of J' are the primes generated by any 3 of the four variables. Since Q_1 is not in any of these primes, we see that $J' + (Q_1)$, and thus also $J' + I$, has codimension 4. It follows by Theorem 18.15c that R is a product of normal domains, and since R is graded with $R_0 = k$ a field, R must itself be a normal domain. Thus we have shown:

The ideal $I = (Q_1, Q_2)$ has codimension 2 iff not all the a_i are equal; it is prime iff no two of the a_i are equal; and in this case it is the ideal of a smooth curve in $\mathbf{P}^3$.

18.4 Flatness and Depth

Most interesting Noetherian rings can be written as finitely generated modules over regular subrings. For example, Noether normalization allows us to write any affine algebra as a finitely generated module over a polynomial ring of the same dimension, and every complete local ring may be written as a finitely generated module over a regular local ring of the same dimension.

The view of a ring as a finitely generated module over a regular subring clarifies many structures. In particular, it yields a dramatic clarification of the nature of the Cohen-Macaulay property. We begin with a more general formulation, based on the local criterion of flatness (Theorem 6.8). The second part provides a converse to Theorem 10.10 in an interesting special case.

Theorem 18.16. *Let (R, P) be a regular local ring, and let (A, Q) be a local Noetherian R-algebra, with $PA \subset Q$.*

a. *A is flat over R iff* $\operatorname{depth}(PA, A) = \dim R$

b. *If A is Cohen-Macaulay, then A is flat over R iff* $\dim A = \dim R + \dim A/PA$.

Proof.

a. Let $\dim R = d$. Since R is regular, P is generated by a regular sequence $x_1, \ldots, x_d$. Let $K = K(x_1, \ldots, x_d)$ be the Koszul complex, which is exact except for $H^d K = k$. If A is flat, then tensoring with A takes exact sequences to exact sequences, so $H^i(K \otimes A) = 0$ for $i \neq d$ too. But $K \otimes A$ may be identified as the Koszul complex over A of the images of the x_i in A, so that these form a regular sequence in A, and PA has depth d.

   Conversely, if $\operatorname{depth}(PA, A) = d$ then the Kozul complex $K \otimes A$ of the generators of PA has no homology except for H^d. Thinking of K as a free resolution of R/P over R, we see that $\operatorname{Tor}_1^R(R/P, A) = H^{d-1}(K \otimes A) = 0$, so A is flat by Theorem 6.8, applied with $S = M = A$.

b. If A is flat over R, then $\dim A = \dim R + \dim A/PA$ by Theorem 10.10 independently of the Cohen-Macaulay property. Conversely, suppose that A is Cohen-Macaulay and $\dim A = \dim R + \dim A/PA$. It follows that $\operatorname{codim} PA = \dim R$. Since $\operatorname{depth}(PA, A) = \operatorname{codim} PA$, we may apply part a and conclude that A is flat over R. □

For a striking special case of this result, see Exercise 18.18.

In case A is finitely generated as an R-module, flatness is the same thing as freeness by Corollary 6.6. In this case we can weaken the hypothesis that A is local.

Corollary 18.17. *Let (R, P) be a regular local ring, and suppose that R is contained in a ring A in such a way that A is a finitely generated R-module. If A is equidimensional, in the sense that the localizations of A at maximal ideals all have the same dimension, then A is Cohen-Macaulay iff A is a free R-module.*

Proof. Again, let $\dim R = d$. Since R is regular, P is generated by a regular sequence $x_1, \ldots, x_d$. We apply the local criterion of flatness, Theorem 6.8, with $S = R$ and $M = A$, and conclude as in the preceding theorem that A is a flat (and thus free) R-module iff $x_1, \ldots, x_d$ is a regular sequence in A.

Suppose that this condition is satisfied. By Proposition 9.2 every maximal ideal Q of A contains the maximal ideal of R, so $x_1, \ldots, x_d$ is a regular sequence in A_Q, and A_Q, is thus Cohen-Macaulay; it follows that A is Cohen-Macaulay too.

Conversely, if A is Cohen-Macaulay, then since $\dim PA = 0$, by Proposition 9.2 we have $\operatorname{codim} PA = d$, and we see that $\operatorname{depth}(P, A) = d$. Since R is local, Corollary 17.7 shows that $x_1, \ldots, x_d$ is a regular sequence on A. □

Corollary 18.17 may be interpreted geometrically as saying (for example, in the affine case) that a variety X is Cohen-Macaulay (that is, has coordinate ring that is Cohen-Macaulay) iff for some (respectively any) finite map from X to a regular variety Y, the fibers of the map all have the same length. See Exercise 18.17.

The Noether normalization theorem in the complete case shows that any complete local ring is a finite module over a regular local subring (see Exercise 13.9 for rings containing a field, and Bourbaki [1983] in general). Therefore Corollary 18.17 can be applied to the completion of any ring. Since a local ring is Cohen-Macaulay iff its completion is Cohen-Macaulay, this gives a reasonably satisfying result in the general case.

Cohen-Macaulay rings have many desirable properties, for example, duality; see Hartshorne [1977, Chapter III] for the beginning of this story. See also Bruns and Herzog [1993] for an idea of the richness of the current theory. In the remainder of this chapter we shall focus on examples.

18.5 Some Examples

From the preceding results we see that any **regular** ring—that is, one whose localizations are regular local rings—is Cohen-Macaulay. The easiest way to generate further examples is to use Proposition 18.13. The importance of such examples leads us to make the following definition.

We say that a ring R is a **complete intersection** if there is a regular ring S and a regular sequence $x_1, \ldots, x_n \in S$ such that $R \cong S/(x_1, \ldots, x_n)$. R is **locally a complete intersection** if this is true for R_P for every maximal ideal P of R. By Propositions 18.8 and 18.13, any ring that is locally a complete intersection is Cohen-Macaulay. As we shall show in the next chapter, the localization of a regular local ring is again regular, and it follows that the same is true for complete intersections—see Exercise 19.2.

It can be shown that if R is a local complete intersection, then for any surjection $S \to R$ from a regular local ring, the kernel is generated by a regular sequence in S, so the definition is actually independent of the regular ring chosen (see, for example, Avramov, Foxby, and Herzog [1994]). Since every complete local ring is a homomorphic image of a regular local ring (see Theorem 7.17 for rings containing a field), it is convenient to weaken the definition of a complete intersection and say that a local ring R is a complete intersection if $\hat{R}$ can be written as a regular local ring modulo a regular sequence. If R itself is a homomorphic image of a regular local ring—and thus in virtually all cases of geometric interest—this apparently weaker definition is equivalent to the previous one. Though it is far from obvious, this more general notion of complete intersection also localizes; see Avramov [1975].

Many interesting rings that are not complete intersections are still Cohen-Macaulay. Here is a short list of examples:

1. **Determinantal rings.** People use this phrase in different ways, but one common usage that suits our needs is the following: A ring R is determinantal if it can be written in the form $R = S/I$ where S is a Cohen-Macaulay ring and I is the ideal generated by the $r \times r$ minors of a $p \times q$ matrix M, for some p, q, r, such that the codimension of I in S is exactly $(p - r + 1)(q - r + 1)$. (Recall from Exercises 10.9 and 10.10 that this is an upper bound for the codimension of I if $I \neq R$; and that if S is a polynomial ring in pq indeterminates, and M is the "generic" matrix with distinct indeterminates for entries, then the condition is satisfied.) Note that complete intersections are determinantal rings with $p = r = 1$. Proposition 18.13 generalizes:

Theorem 18.18. *Determinantal rings are Cohen-Macaulay.*

This result was proved by J. Eagon in the special case $r = \min(p, q)$ and by Eagon and Hochster in general. See Bruns and Vetter [1988] for a recent treatment that includes a simple proof.

Of course, an unmixedness theorem for determinantal ideals generalizing Corollary 18.14 follows. Interestingly this was proved by Macaulay himself (for the maximal minors), and his work was what led to the whole development we have described.

2. Similarly, one can make rings using the minors of a symmetric matrix, the "Pfaffians" of a skew-symmetric matrix, and so on. Also interesting are the ideals defining the varieties of square matrices φ such that the rank of $\varphi^r \leq d_r$ for some sequence of integers d_r. For each of these ideals a result similar to Theorem 18.18 holds, with different codimensions required, depending on the case. See Bruns and Vetter [1988], De Concini, Eisenbud and Procesi [1982], and Eisenbud and Saltman [1989] for these and related results.

3. **Invariants.** Let S be a polynomial ring over a field k, and let G be a linearly reductive algebraic group, acting algebraically by linear transformations of the variables of S. The ring of invariants S^G, consisting of all those elements fixed by G, is an interesting object. We saw in Chapter 1 that S^G is a finitely generated k-algebra (Hilbert's theorem). Hochster and Roberts have shown that S^G is also Cohen-Macaulay. In the case of a finite group G this is fairly easy—see Exercise 18.14. The initial proofs in the general case were quite difficult, but recently Hochster and Huneke's technique of **tight closure** [1990] has given rise to a much simpler proof; see Bruns and Herzog [1993, Section 6.5]. Again, we shall not give any proof; rather, we describe a few examples.

   A finite group G is linearly reductive as long as the characteristic of k does not divide the order of G. The classical groups $\mathrm{GL}(n)$, $\mathrm{SL}(n)$,

O(n), etc., are linearly reductive in characteristic 0. The algebraic torus consisting of a product of copies of the multiplicative group (of k, if k is algebraically closed, or in the sense of schemes) is linearly reductive in any characteristic. A consequence of the last case, for example, is that if $T \subset \mathbf{Z}_+^n$ is a "rational polyhedral convex cone" (that is, the set of all lattice points in some convex region closed under multiplication by positive numbers, with finitely many sides, having rational slopes, such as the shaded region in Figure 18.3), then $k[T] \subset k[x_1, \ldots, x_n]$ is a Cohen-Macaulay subring. Equivalently, as was proved in Exercise 4.23, these are the normal subrings of $k[x_1, \ldots, x_n]$ generated by finitely many monomials. Fulton [1992] is an excellent introduction to these rings and the varieties made from them.

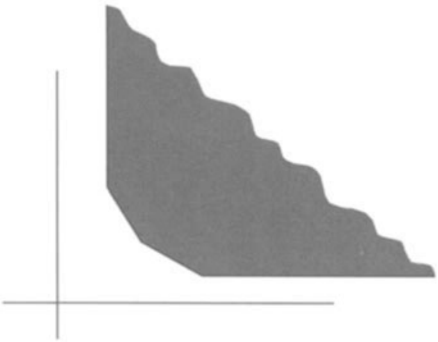


FIGURE 18.3.

Another particularly interesting example is to take S to be the polynomial ring on $\binom{n}{r}$ variables, thought of as a basis for the vector space $\wedge^r k^n$, and take the invariants under the action of $GL(n)$, acting in the natural way on k^n and thus on $\wedge^r k^n$. The invariant ring $R = S^{GL(n)}$ is in this case the homogeneous coordinate ring of the Grassmann variety of r-planes in k^n, so this coordinate ring is Cohen-Macaulay.

4. Let C be a nonsingular algebraic curve of genus g, and let C be embedded in some projective space $\mathbf{P}^r$ by a complete linear series of degree $d \geq 2g$ (or by the complete canonical series). Then the homogeneous coordinate ring of C is Cohen-Macaulay; see Saint-Donat [1973].

   In general, there is a natural characterization of Cohen-Macaulayness for the homogeneous coordinate rings of projective varieties in terms of sheaf cohomology (see Exercise 18.16).

The richness of the Cohen-Macaulay theory has also influenced the theory of arbitrary Noetherian rings, through Hochster's theorem that any local

Noetherian ring containing a field posesses a module (which may not be finitely generated) whose depth is equal to the dimension of the ring—a "maximal Cohen-Macaulay module". See Hochster [1975] for this and related matters.

18.6 Exercises

Throughout these exercises all rings considered are assumed to be Noetherian.

Exercise 18.1: Characterize Cohen-Macaulay rings among 1-dimensional rings in terms of the associated primes of the ring.

Exercise 18.2: Show that a local ring R is regular iff the maximal ideal of R can be generated by a regular sequence.

Exercise 18.3: Give an elementary proof of Lemma 18.3 using Nakayama's lemma and the theory of associated primes.

Exercise 18.4: Let R be a ring and let M be a finitely generated R-module. For any ideal I of R, set

$$\operatorname{codim}(I, M) = \operatorname{codim}_{R/\operatorname{ann}(M)}(I + \operatorname{ann}(M)/\operatorname{ann}(M)).$$

Prove the following generalization of Theorem 18.7: If for every maximal ideal P of R containing $\operatorname{ann}(M)$ we have $\operatorname{depth}(P, M) = \operatorname{codim}(P, M)$, then for every ideal I of R we have $\operatorname{codim}(I, M) = \operatorname{depth}(I, M)$.

Exercise 18.5: Let (R, P) be a local ring, and let M be a finitely generated R-module. If $\operatorname{depth}(P, M) = \dim M$, show that all the associated primes of M have the same dimension, namely $\dim M$.

Exercise 18.6: Prove that the direct product of two Cohen-Macaulay rings is Cohen-Macaulay. Use Exercise 2.25 to show that an affine ring that is Cohen-Macaulay is a product of equidimensional Cohen-Macaulay rings. (The corresponding geometric statement is clear from Corollary 18.11.)

Exercise 18.7:* Prove that the homogeneous coordinate ring of the "twisted cubic,"

$$R = k[s^3, s^2t, st^2, t^3] \subset k[s, t],$$

is Cohen-Macaulay when localized at the maximal homogeneous ideal by finding an explicit homogeneous regular sequence. (We shall show in Exercise 19.10 that if a positively graded ring is Cohen-Macaulay locally at the maximal homogeneous ideal, then it is Cohen-Macaulay.) Check that R is a determinantal ring, by showing that it is isomorphic to $k[x_0, \ldots, x_3]/I$, where I is the ideal generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{pmatrix}.$$

Exercise 18.8:* Prove that the homogeneous coordinate ring of a smooth rational quartic in three-space,

$$R = k[s^4, s^3t, st^3, t^4] \subset k[s,t]$$

is not Cohen-Macaulay. Check that R is isomorphic to the ring $k[x_0, \dots, x_3]/I$, where I is the ideal generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x_0 & x_1^2 & x_1x_3 & x_2 \\ x_1 & x_0x_2 & x_2^2 & x_3 \end{pmatrix}$$

(Actually I is already generated by four of the six minors: One of the cubics and the quartic are superfluous.)

The difference between these two examples may be accounted for as follows: First, the second represents a curve embedded by an incomplete linear series; quite generally, the homogeneous coordinate ring of a variety in an incomplete embedding is never Cohen-Macaulay. (This is best proved with local cohomology, Appendix 4.) Second, as mentioned in the text, ideals generated by minors of a matrix in a Cohen-Macaulay ring define Cohen-Macaulay factor rings if they have the "generic" codimension; for the 2×2 minors of a 2×3 matrix this codimension is 2, and for the 2×2 minors of a 2×4 matrix it is 3, whereas the ideals I and J in the preceding two exercises both have codimension 2.

Exercise 18.9: Deduce from Exercise 11.10 that a Cohen-Macaulay ring is reduced iff it is generically reduced—that is, iff its localization at each minimal prime is reduced. Give an example of a generically reduced ring that is not reduced.

Exercise 18.10 (Max Noether's *AF*+*BG* theorem): The following is a precursor of the unmixedness theorem; prove it as an application. Suppose that F and G are homogeneous forms in three variables generating the ideals of plane curves that meet transversely in a set of points Γ. If H is homogeneous form vanishing on Γ, then there are homogeneous forms A and B such that $H = AF + BG$.

Exercise 18.11 (Eisenstein's Criterion):* Let R be a domain, and let P be a prime ideal of R. If a polynomial $f(x) = r_0x^n + r_1x^{n-1} + \cdots + r_n$ satisfies the conditions $r_0 \notin P, r_1, \dots, r_n \in P, r_n \notin P^2$, then f is irreducible.

Exercise 18.12: Let k be an algebraically closed field of characteristic not 2, and let S be a polynomial ring over k in n variables. Any quadric Q takes on a diagonal form $Q = x_1^2 + x_2^2 + \cdots + x_r^2$ with respect to a suitable choice of coordinates $x_1, \dots, x_n$. (*Proof*: Take an orthonormal basis for the

associated bilinear form modulo its null space.) The number of squares, r, is called the **rank** of Q.

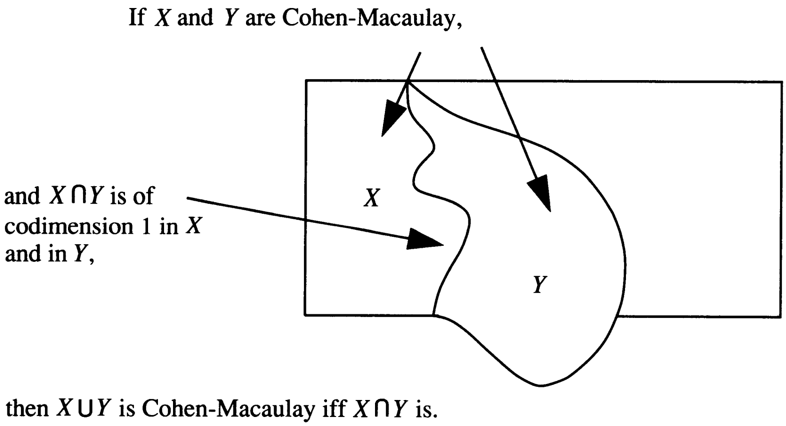
a. Use the Jacobian criterion to find the singular locus of Q. What is its dimension in terms of rank Q?

b. Assuming Q is nonzero, use Theorem 18.15 to show that $S/(Q)$ is reduced iff $r \geq 2$; a domain iff $r \geq 3$; normal iff $r \geq 4$; and corresponds to a smooth projective variety iff $r = n$.

c. For those who know about quadratic forms: What is the corresponding result in characteristic 2? (See, for example, Buchweitz, Eisenbud and Herzog [1987] for further information.)

Exercise 18.13: Let I, J be ideals of a local ring R such that $I \cap J = 0$. Suppose that $R/I, R/J$, are Cohen-Macaulay rings of the same dimension d, and that $R/(I+J)$ is of dimension $d-1$. Show that R is Cohen-Macaulay iff $R/(I+J)$ is.

An affine algebraic set X is called Cohen-Macaulay if $A(X)$ is Cohen-Macaulay. In these terms we can reformulate the result of this exercise by saying: If X and Y are Cohen-Macaulay, and $X \cap Y$ is of codimension 1 in X and in Y, then $X \cup Y$ is Cohen-Macaulay iff $X \cap Y$ is.



Exercise 18.14: Let G be a finite group acting as automorphisms of an algebra R over a field of characteristic 0. Show that if R is Cohen-Macaulay, then the ring of invariants R^G is Cohen-Macaulay.

Exercise 18.15: Let $\Gamma \subset \mathbf{P}^r$ be a finite set of points. Show that the homogeneous coordinate ring of Γ is Cohen-Macaulay.

Exercise 18.16: (For those with knowledge of the cohomology of coherent sheaves) Prove:

a. The homogeneous coordinate ring of a curve $X \subset \mathbf{P}^r$ with ideal sheaf I is Cohen-Macaulay iff $H^1(\mathbf{P}^r; I(n)) = 0$ for all n iff X is linearly normal and $\oplus_n H^0(\mathbf{P}^r; \mathcal{O}_X(n))$ is generated as an algebra by $H^0(\mathbf{P}^r; \mathcal{O}_X(1))$.

b. The homogeneous coordinate ring of a projective variety $X \subset \mathbf{P}^r$ of pure dimension d with ideal sheaf I is Cohen-Macaulay iff $H^1(\mathbf{P}^r; I(n)) = 0$ for all n and $H^i(\mathbf{P}^r; \mathcal{O}_X(n)) = 0$ for $0 < i < d$, and all n.

Exercise 18.17 (Cohen-Macaulayness in the geometric case): Here is the most commonly used geometric form of Corollary 18.17: Let X be a projective algebraic set whose components all have dimension d. If $\pi : X \to Y$ is a surjective morphism onto a locally regular variety of dimension d, then X is locally Cohen-Macaulay iff π is flat iff all the fibers $\mathcal{O}_{X,p}/\mathfrak{m}_{Y,\pi(p)}$ have the same length. (You may need to use Exercise 20.13.)

Exercise 18.18 (Hartshorne, [1966a]):* Suppose (R, P) is a local ring containing a field k, and let $x_1, \ldots, x_r \in P$ be a sequence of elements. Show that $x_1, \ldots, x_r$ is a regular sequence iff R is flat as a module over $k[x_1, \ldots, x_r]$.

19

Homological Theory of Regular Local Rings

In this section we shall examine some further uses of the homological tools. After preliminaries on minimal free resolutions, we shall use the Koszul complex to prove the Hilbert syzygy theorem and the basic results of Auslander-Buchsbaum and Serre about regular local rings and polynomial rings: the characterization of regular local rings as the rings of finite global dimension, and the consequences that localizations of regular local rings are regular and that regular local rings are factorial. We also derive some important relations of depth to homology in the Auslander-Buchsbaum formula and the formula connecting the vanishing of $\operatorname{Ext}^i_R(M,N)$ with the depth of the annihilator of M on N. This "explains" in a certain sense the relations of depth and the homology of the Koszul complex that we saw in Chapter 18.

19.1 Projective Dimension and Minimal Resolutions

We begin with some basic ideas.

Definition. *A* ***projective resolution*** *of an R-module M is a complex*

$$\mathcal{F}: \ \cdots \to F_n \xrightarrow{\varphi_n} \cdots \to F_1 \xrightarrow{\varphi_1} F_0$$

of projective R-modules such that coker $\varphi_1 = M$ and $\mathcal{F}$ has no homology. (Sometimes we add a "$\to 0$" to the right-hand side of $\mathcal{F}$, and then insist that

$\mathcal{F}$ have no homology except at F_0.) We shall sometimes abuse this notation and say that

$$\mathcal{F}: \cdots \to F_n \xrightarrow{\varphi_n} \cdots \to F_1 \xrightarrow{\varphi_1} F_0 \to M \to 0$$

*is a resolution of M. $\mathcal{F}$ is a **free resolution** if all the F_i are free, and a **graded free resolution** if R is a graded ring, all the F_i are graded free modules (that is, sums of modules of the form $R(d)$, for various integers d), and the maps are all homogeneous maps of degree 0 (that is, they take homogeneous elements to homogeneous elements of the same degree). Of course, only graded modules can have graded free resolutions. If for some $n < \infty$ we have $F_{n+1} = 0$, but $F_i \neq 0$ for $0 \leq i \leq n$, then we shall say that $\mathcal{F}$ is a **finite** resolution, of **length** n.*

It is easy to see that every module has a free resolution and, if R is graded, every graded module has a homogeneous free resolution. To construct one, begin by taking a set of generators for M and map a free module onto M sending the free generators of the free module to the given generators of M. Let M_1 be the kernel of this map, and repeat the procedure, now starting with M_1.

Free or projective resolutions serve to compare a module with free modules. As we have explained in Chapter 1, free resolutions were originally studied by Hilbert [1890] in the case of a graded module M over a polynomial ring $R = k[x_0, \ldots, x_r]$, in order to compute the **Hilbert function** of the module M—the function

$$H_M(n) = \dim_k M_n.$$

For this one needs a finite free resolution, and we proved in Chapter 15 that every module over a polynomial ring has a finite free resolution. In this chapter we shall extend this result to regular local rings and derive some of its consequences.

In general, we define the **projective dimension** of M, written $\operatorname{pd} M$ (or $\operatorname{pd}_R M$ if the ring involved is not clear from context), to be the minimum of the lengths of projective resolutions of M (it is ∞ if M has no finite projective resolution). The **global dimension** of R is the supremum of the projective dimensions of all R-modules. The following result of Auslander shows that it is enough to take the supremum for finitely generated R-modules.

Theorem 19.1 (Auslander [1955]). *The following conditions on a ring R are equivalent:*

a. $\operatorname{gl\,dim} R \leq n$*—that is,* $\operatorname{pd} M \leq n$ *for every R-module M.*

b. $\operatorname{pd} M \leq n$ *for every finitely generated module M.*

A proof (together with some other equivalent conditions) is given in Theorem A3.18. We shall not actually use the result in this book—for our

purposes we might as well have defined the global dimension to be the supremum of the projective dimensions of finitely generated modules. In particular, we shall state and prove the results of this chapter for finitely generated modules, although many of them hold in the case of arbitrary modules as well.

There is a simplification in the local and graded cases that is essentially a consequence of Nakayama's lemma: The notions of projective and free modules coincide. This gives rise to a characterization of projectives as "locally free" modules that we shall need later.

Theorem 19.2 (Characterization of projectives). *Let M be a finitely generated module over a Noetherian ring R. The following statements are equivalent:*

a. M is a projective module.

b. M_P is a free module for every maximal ideal (and thus for every prime ideal) P of R.

c. There is a finite set of elements $x_1, \ldots, x_r \in R$ that generate the unit ideal of R such that $M[x_i^{-1}]$ is free over $R[x_i^{-1}]$ for each i.

In particular, every projective module over a local ring is free. Every graded projective module over a positively graded ring R with R_0 a field is a graded free module.

A proof of Theorem 19.2 is contained in Exercises 4.11 and 4.12, and the hints provided there. The result is "responsible" for the fact that one may identify projective modules and vector bundles; see Corollary A3.3 for a sketch. The characterization of projectives as locally free modules is also true without the Noetherian and finitely generated hypotheses. See Kaplansky [1958] for the general case.

We have already studied one family of finite free resolutions in some detail: the Koszul complexes of regular sequences. They yield:

Corollary 19.3. *If $x = x_1, \ldots, x_n$ is a regular sequence, then $K(x)$ is a free resolution of $R/(x_1, \ldots, x_n)$. In particular, if R is a regular local ring, and $x_1, \ldots, x_n$ is a minimal set of generators for the maximal ideal of R, then the Koszul complex $K(x_1, \ldots, x_n)$ is a finite free resolution of the residue class field of R.*

Proof. The first statement is a restatement of Corollary 17.5 in the case $M = R$. That the generators of the maximal ideal in a regular local ring form a regular sequence was proved in Corollary 10.15. □

The surprising fact is that Corollary 19.3 suffices to show that every finitely generated R-module has a free resolution of length at most n. (With Theorem 19.1, this even implies that every module has a free resolution of length at most n.) Before deriving this implication, we digress to discuss resolutions over local rings more generally.

Rather than starting a free resolution of a module M by selecting an arbitrary set of generators of M, it seems reasonable to think we would do better to choose a minimal set of generators. (One might also try a set of generators with some other special property—as we did in the context of Gröbner bases in Chapter 15.) Obtaining in this way a "minimal" map from a free module F to M, we could continue by choosing a minimal set of generators for the kernel, and so forth, and get a "minimal" free resolution. It turns out that this idea is not useful in general, since minimal sets of generators have no uniqueness properties. (Even the number of generators in a minimal set of generators need not be well defined; for example, in the integers, $(5) = (10,15)$.) But over a local or graded ring, where Nakayama's lemma makes the notion of a minimal set of generators nice, the idea comes fully into its own; as we shall see in Chapter 20, each module over a local ring has a unique minimal free resolution, and every resolution is constructed from the minimal one in a simple way.

To avoid saying things twice, we shall work only with the local case; but the reader will have no difficulty in translating everything for the case of a positively graded ring with degree 0 part a field. We shall use both in the sequel.

It turns out to be useful to define not only the notion of a minimal free resolution, but the notion of a minimal complex as well, and for this the definition has to be cast in a form apparently different from the previous one.

Definition. *A complex*

$$\mathcal{F}: \quad \cdots \to F_n \xrightarrow{\varphi_n} F_{n-1} \to \cdots$$

*over a local ring (R, P) is **minimal** if the maps in the complex $\mathcal{F} \otimes R/P$ are all 0; that is, for each n, the image of $\varphi_n : F_n \to F_{n-1}$ is contained in PF_{n-1}.*

In case $\mathcal{F}$ is a complex of free modules, this simply means that any matrix representing φ_n has all its entries in P.

For example, if $x := (x_1, \ldots, x_n) \in R^n$ and each component x_i is in P, then the Koszul complex $K(x)$ is a minimal complex since the maps in the complex $K(x) \otimes R/P$ are given by the exterior product with the image of the element x in R^n/PR^n—and this image is 0.

The relation of this peculiar definition to minimality in the previous sense is given by the following lemma.

Lemma 19.4. *A free resolution*

$$\mathcal{F}: \cdots \to F_n \xrightarrow{\varphi_n} F_{n-1} \to \cdots \xrightarrow{\varphi_1} F_0$$

over a local ring is a minimal complex iff for each n*, a basis of* F_{n-1} *maps onto a minimal set of generators of coker* φ_n.

Proof. Let R, P be the local ring, and let φ_0 be the natural map $F_0 \to$ coker φ_1. For any $n \geq 0$, consider the induced epimorphism of vector spaces

$$F_{n-1}/PF_{n-1} \twoheadrightarrow (\operatorname{coker}\varphi_n)/P(\operatorname{coker}\varphi_n).$$

By Nakayama's lemma, a basis for the vector space on the right is a minimal set of generators of coker φ_n. Thus the second condition of the lemma is satisfied iff this epimorphism is an isomorphism. This is equivalent to the condition that im φ_n is in PF_n, which is the condition of minimality. □

The following useful consequence might be described as a homological version of Nakayama's lemma.

Corollary 19.5. *If* R *is a local ring with residue class field* k*, and* M *is a finitely generated nonzero* R*-module, then* $\operatorname{pd}_R M$ *is the length of every minimal free resolution of* M*. Further,* $\operatorname{pd}_R M$ *is the smallest integer* i *for which* $\operatorname{Tor}_{i+1}^R(k, M) = 0$*. Thus the global dimension of* R *is equal to* $\operatorname{pd}_R k$.

The proof of the last statement rests on the fact that if R is a ring and M and N are R-modules, then the module $\operatorname{Tor}_i^R(M, N)$ can be computed either as the ith homology of the tensor product of M with a projective resolution of N. See Theorem A3.24, Application i.

Proof. $\operatorname{Tor}_{i+1}^R(k, M)$ can be computed as the $(i+1)$ homology module of the tensor product of k and an arbitrary resolution of M. Thus if $n = \operatorname{pd}_R M$, then $\operatorname{Tor}_{i+1}^R(k, M) = 0$ for $i \geq n$.

Now suppose that

$$\mathcal{F}: 0 \to F_n \to F_{n-1} \to \cdots \to F_0$$

is a free resolution of M of length n. If i is the smallest integer for which $\operatorname{Tor}_{i+1}^R(k, M) = 0$, then we trivially have $n \geq \operatorname{pd}_R M \geq i$. But if $\mathcal{F}$ is minimal then the differentials in the complex $k \otimes \mathcal{F}$ are 0, so

$$\operatorname{Tor}_{i+1}^R(k, M) = k \otimes F_{i+1}.$$

This is 0 iff F_{i+1} is 0, so $i = n$, proving the first two statements of the corollary.

Since we may compute $\operatorname{Tor}_{i+1}^R(k, M)$ from a free resolution for k as well, $\operatorname{Tor}_{i+1}^R(k, M) = 0$ for $i \geq \operatorname{pd}_R k$, and we see that $\operatorname{pd}_R M \leq \operatorname{pd}_R k$ for any finitely generated R-module. Combining this with Auslander's theorem (19.1) we get the last statement. □

19.2 Global Dimension and the Syzygy Theorem

We return at last to regular local rings.

Corollary 19.6. *If R is a regular local ring of dimension n, then the global dimension of R is n.*

We shall see in Theorem 19.12 that regular local rings are the only local rings with finite global dimension.

Proof. If $x_1, \ldots, x_n$ generate the maximal ideal of R, then we have seen that the Koszul complex $K(x_1, \ldots, x_n)$ is a minimal free resolution of length n of the residue class field k of R. By Corollary 19.5, $n = \operatorname{pd}_R k$ is equal to the global dimension of R.

As we have already remarked, the case of graded modules over a positively graded ring is completely analogous to the local case. The only feature of regularity that we just used is that the maximal ideal is generated by a regular sequence. This condition is satisfied by the ideal of positively graded elements of a graded polynomial ring, so imitating the preceding proof, we get another proof of Theorem 1.13, which is opposite in spirit from the one contained in Corollary 15.11.

Corollary 19.7 (Hilbert Syzygy Theorem). *If k is a field, then every finitely generated graded module over $k[x_1, \ldots, x_n]$ has a graded free resolution of length $\leq n$.*

The proof of the syzygy theorem that we have presented is due to Cartan and Eilenberg [1956]. It looks like a bit of homological trickery compared with Hilbert's original proof. Hilbert was strongly influenced by the elimination theory of his time, and his proof is a constructive reduction to a problem over a polynomial ring with fewer variables, closer in spirit to the proof of Corollary 15.13.

Somewhat surprisingly, Corollary 19.6 implies the corresponding result in the ungraded case.

Corollary 19.8. *Every finitely generated module over $k[x_1, \ldots, x_n]$ has a finite free resolution.*

Proof. Let $F \xrightarrow{\varphi} G \to M \to 0$ be a free presentation of the finitely generated module M over the polynomial ring $S = k[x_1, \ldots, x_n]$. Introducing a new variable x_0, we may "homogenize φ" as follows: First choose bases, so that φ is represented by a matrix, and let d be the maximum of the degrees of the polynomials appearing in this matrix. Next, replace each monomial in each entry of φ by that monomial multiplied by the power of x_0 necessary

to bring its degree up to d; let $\tilde{\varphi}$ be the matrix over $T = k[x_0, x_1, \dots, x_n]$ whose entries are the resulting homogeneous polynomials of degree d.

Note that we may write $S \cong T/(1 - x_0)$. With this module structure on S, we claim that $\varphi = \tilde{\varphi} \otimes_T S$; indeed, this equation simply says that if we replace x_0 by 1 in each entry of $\tilde{\varphi}$, we get back φ, which is obvious from the construction. If we let $M^{\sim} = \operatorname{coker} \tilde{\varphi}$, we thus have $M^{\sim} \otimes_T S = M$.

Now let $\mathcal{F}^{\sim}$ be a free resolution of $M^{\sim}$, beginning with $\tilde{\varphi}$, which exists by virtue of the Hilbert syzygy theorem (applied, for instance, to $\ker \tilde{\varphi}$). We shall complete the proof by showing that $\mathcal{F}^{\sim} \otimes_T S$ is a resolution of M. For this it suffices to show that $\mathcal{F}^{\sim} \otimes_T S$ has no homology (except M, at the zeroeth step), that is, that $\operatorname{Tor}_i^T(M^{\sim}, S) = 0$ for $i \geq 1$. We may compute this module from the free resolution

$$0 \to T \xrightarrow{1-x_0} T \to S \to 0$$

of S; tensoring with $M^{\sim}$, we see we must show that

$$0 \to M^{\sim} \xrightarrow{1-x_0} M^{\sim}$$

is exact, that is, that $1 - x_0$ is a nonzerodivisor on $M^{\sim}$. But $1 - x_0$ is a nonzerodivisor on any graded module M, since for any element

$$m = m_e + (\text{degree greater than } e) \in M,$$

with $\deg m_e = e$, we have

$$(1 - x_0)m = m_e + (\text{degree greater than } e). \qquad \square$$

For the generalization to polynomial rings over regular rings, see Exercise 19.3.

19.3 Depth and Projective Dimension: The Auslander-Buchsbaum Formula

In order to exploit the fact that regular local rings have finite global dimension, we shall use a connection between projective dimension and depth discovered by Auslander and Buchsbaum.

Theorem 19.9 (Auslander-Buchsbaum formula). *Let (R, P) be a local ring. If M is a finitely generated R-module of finite projective dimension, then*

$$\operatorname{pd} M = \operatorname{depth}(P, R) - \operatorname{depth}(P, M).$$

See Exercise 19.8 for the graded case. If R is a regular local ring, this formula follows at once from Corollary 19.5 and Theorem 17.4, because if the maximal ideal of R is generated by the regular sequence $x_1, \dots, x_n$, we have

$$\mathrm{Tor}_{i+1}^R(k, M) = H^{n-i-1}(M \otimes K(x_1, \ldots, x_n)).$$

The left side computes the projective dimension while the right side computes the depth.

Proof. We exploit the finiteness of the projective dimension of M by using induction; if $\mathrm{pd}\, M = 0$ then M is free and the result is obvious.

If $\mathrm{pd}\, M > 0$, let

$$\mathcal{F} : 0 \to N \xrightarrow{\varphi} F \to M \to 0$$

be one step of a minimal free resolution of M—that is, let $F \to M$ be an epimorphism from a free module F of minimum possible rank, and let N be the kernel, with $\varphi : N \to F$ the inclusion. By Corollary 19.5 we have $\mathrm{pd}\, N = \mathrm{pd}\, M - 1$, so we may apply the theorem inductively to N. Writing $d = \mathrm{depth}\ N$, we must show that the depth of M is exactly $d - 1$.

To do this we shall exploit the characterization of depth by the Koszul complex. Let $x = (x_1, \ldots, x_n)$ be a set of generators of the maximal ideal. If we tensor the Koszul complex $K(x)$ with the short exact sequence $\mathcal{F}$, we obtain a short exact sequence of complexes which gives the following long exact homology sequence:

$$\begin{aligned} \cdots \to H^{i-1}(F \otimes K(x)) &\to H^{i-1}(M \otimes K(x)) \to H^i(N \otimes K(x)) \\ \to H^i(F \otimes K(x)) &\to \cdots . \end{aligned}$$

Since N and F both have depth $\geq d$, we see at once that

$$H^i(M \otimes K(x)) = 0$$

for $i < d - 1$. To prove that depth $M = d - 1$ it therefore suffices to show that

$$H^{d-1}(M \otimes K(x)) \neq 0.$$

Since depth $N = d$ we know $H^d(N \otimes K(x)) \neq 0$. It is thus more than sufficient to prove that the map

$$H^d(N \otimes K(x)) \to H^d(F \otimes K(x)),$$

which is the map induced by φ, is zero.

If $\mathrm{pd}\, N > 0$, then by the theorem applied to N, we have $d < \mathrm{depth}\, R$, so that in fact $H^d(F \otimes K(x)) = 0$. Otherwise, $\mathrm{pd}\, N = 0$, so that N is free, and we have

$$\begin{aligned} H^d(N \otimes K(x)) &= N \otimes H^d(K(x)), \\ H^d(F \otimes K(x)) &= F \otimes H^d(K(x)). \end{aligned}$$

The map induced by φ is in these terms simply $\varphi \otimes 1$. Since φ is a minimal presentation, it may be represented by a matrix with entries in P. But $H^d(K(x))$ is annihilated by P by Proposition 17.14, so the tensor product map is in fact 0, as required.

The Auslander-Buchsbaum formula is a fundamental tool for studying modules of finite projective dimension. Here is a first indication of its usefulness: A result connecting projective dimension to the theory of primary decomposition.

Corollary 19.10. *Let (R, P) be a local ring. If there exists a finitely generated module of projective dimension equal to the dimension of R, then R is Cohen-Macaulay. If R is Cohen-Macaulay, then a module M of finite projective dimension has* $\operatorname{pd} M = \dim R$ *iff the maximal ideal is associated to M.*

We can also deduce a result of Bayer-Stillman comparing the projective dimension of an ideal I and its generic initial ideal $\operatorname{Gin}(I)$, using Theorem 15.13. A similar result, for the power series case, was announced by Grauert in [1972]:

Corollary 19.11. *Let $S = k[x_1, \ldots, x_r]$ be the polynomial ring over an infinite field, and let F be a graded free S-module with given basis and a reverse lexicographic monomial order. If $I \subset S$ is a homogeneous ideal, then*

$$\operatorname{pd}_S S/I = \operatorname{pd}_S S/\operatorname{Gin}(I).$$

Proof. We use the graded version of the Auslander-Buchsbaum Theorem, Exercise 19.8. If $\operatorname{depth}((x_1, \ldots, x_r), S/I) \geq 1$, then by prime avoidance some linear form is a nonzerodivisor on S/I. After a generic change of coordinates, we may assume that x_r is a nonzerodivisor on S/I in this case. Continuing in this way, we may suppose, after a generic change of coordinates, that $x_r, x_{r-1}, \ldots, x_s$ is a maximal S/I-regular sequence. By Theorem 15.20, $\operatorname{Gin}(I)$ is Borel-fixed, so Corollary 15.25 shows that there is a maximal $(S/\operatorname{Gin}(I))$-regular sequence of the form $x_r, x_{r-1}, \ldots, x_t$. By Theorem 15.13, $s = t$, so $\operatorname{depth} S/I = \operatorname{depth} S/\operatorname{Gin}(I)$, and we are done. □

A similar result holds for arbitrary modules; we leave to the reader the necessary generalizations.

One of the most significant applications of the Auslander-Buchsbaum formula is the promised completion of Corollary 19.6, a fundamental result of Auslander-Buchsbaum and Serre:

Theorem 19.12. *A local ring has finite global dimension iff it is regular.*

Proof. Half of this is done by Corollary 19.6. We now must show that if R has finite global dimension then R is regular.

Suppose that R has finite global dimension, and let k be its residue class field. Let $x_1, \ldots, x_n$ be a minimal set of generators of the maximal ideal of R.

We must show $\dim R = n$. By the principal ideal theorem (Theorem 10.2) we have $\dim R \leq n$, and it suffices to prove the opposite inequality.

By Proposition 18.2 it suffices to show that depth $R \geq n$. But, by the Auslander-Buchsbaum formula, Theorem 19.9, depth $R = \operatorname{pd} k$. In Lemma 19.13 we shall show that the Koszul complex $K(x_1, \ldots, x_n)$, which has length n, is contained in the minimal free resolution of k. In particular, $\operatorname{pd} k \geq n$, and we are done. □

For a generalization of the fact that if R has finite global dimension then R is regular, see Exercise 20.23.

It remains to prove the following more-general result:

Lemma 19.13. *If (R, P) is a local ring with residue class field k, and if P is minimally generated by $x_1, \ldots, x_n$, then $K(x_1, \ldots, x_n)$ is a subcomplex of the minimal free resolution of k.*

Proof. Let

$$\mathcal{F}: \quad \cdots \to F_1 \to F_0$$

be the minimal free resolution of k. We have trivially a comparison map of complexes $\varphi : K(x_1, \ldots, x_n) \to \mathcal{F}$ lifting the identity map $k \to k$. We shall show by induction on i that, for each i, the map

$$\varphi_i : \wedge^{n-i} R^n \to F_i,$$

is a split monomorphism.

The statement is obvious for $i = 0$ and 1; since in fact

$$(\wedge^{n-1} R^n \cong R^n) \to (\wedge^n R^n \cong R) \to k \to 0$$

is a minimal free presentation of k, it is isomorphic to

$$F_1 \to F_0 \to k \to 0$$

via φ_0 and φ_1.

Suppose, inductively, that we know that the map φ_{i-1} is a split monomorphism. Consider the following diagram.

$$\begin{array}{ccc} F_i & \longrightarrow & F_{i-1} \\ {\scriptstyle \varphi_i}\uparrow & & {\scriptstyle \varphi_{i-1}}\uparrow \\ \wedge^{n-i} R^n & \xrightarrow{d} & \wedge^{n-i+1} R^n \end{array}$$

It will be enough to show that

$$R/P \otimes \varphi_i : R/P \otimes \wedge^{n-i} R^n \to R/P \otimes F_i$$

is a monomorphism, since then by Nakayama's lemma a minimal set of generators of $\wedge^{n-i} R^n$ maps to a subset of a minimal set of generators for F_i.

Note that since the differential d of the Koszul complex maps into the maximal ideal times $\wedge^{n-i+1} R^n$, it induces a map of vector spaces

$$\bar{d} : R/P \otimes \wedge^{n-i} R^n \to P/P^2 \otimes \wedge^{n-i+1} R^n.$$

We shall show that $\bar{d}$ is a monomorphism; since φ_{i-1} is a split monomorphism, it takes $P/P^2 \otimes \wedge^{n-i+1} R^n$ monomorphically to $P/P^2 \otimes F_{i-1}$, and this implies that $R/P \otimes \varphi_i$ is a monomorphism, as desired.

Finally, the proof that $\bar{d}$ is a monomorphism is nothing but linear algebra. Since the elements x_i minimally generate P, the vector space P/P^2 is isomorphic to k^n, with basis $\{x_j\}$. Writing $\{e_j\}$ for a basis of R^n, we wish to show that the map

$$\wedge^{n-i} k^n \to k^n \otimes \wedge^{n-i+1} k^n$$

given by

$$a \to \sum_j x_j \otimes e_j \wedge a$$

is a monomorphism for each $i < n$. Since the elements x_j are linearly independent, it suffices to show that not all the $e_j \wedge a$ can be zero unless a is zero. This follows at once by direct computation, or from the observation that the multiplication map $\wedge^{n-i} k^n \times \wedge^i k^n \to \wedge^n k^n = k$ is a perfect pairing.

The inclusion of complexes in Lemma 19.13 makes the free modules in the Koszul complex into direct summands of the free modules in the minimal free resolution. In fact, an even stronger result holds: Results of Assmus, Tate, and Gulliksen (see, for example, Gulliksen and Levin [1969]) show that the minimal free resolution of k has the structure of a free graded skew-commutative algebra (with the differentials acting as derivations) and that the Koszul complex, regarded as the exterior algebra, is a tensor factor.

As a consequence of Theorem 19.12 we can prove that a localization of a regular local ring is again regular. The result was known originally only for the geometric cases (affine rings, power-series rings, and so forth). The proof of the general case, by Auslander and Buchsbaum and by Serre independently, was one of the first big successes of "representation theory"—using module theory to prove results about rings—in commutative algebra.

Corollary 19.14. *Every localization of a regular local ring is regular. Every localization of a polynomial ring over a field is regular.*

Proof. Suppose R is a regular local ring or a polynomial ring, and let R_P be a localization. By Theorem 19.12 it is enough to show that R_P has finite global dimension, and by Corollary 19.5 it suffices to prove that the residue field R_P/P_P has finite projective dimension. Since R is regular, R/P has a finite projective resolution over R, and the localization of this is a finite free resolution of R_P/P_P over R_P, and we are done.

The Auslander-Buchsbaum formula can also be used to give an "extrinsic" characterization of Cohen-Macaulay rings:

Corollary 19.15. *If R is a regular local ring and A is a local R-algebra that is finitely generated as an R-module, then A is Cohen-Macaulay iff* $\operatorname{pd}_R A = \operatorname{codim}_R A$ (*the codimension of the annihilator of A in R*).

For example, if A is a local ring that is a finitely generated module over a regular local ring $R \subset A$, then Corollary 19.15 implies that A is Cohen-Macaulay iff A is a free R-module (compare Corollary 18.17).

Proof. Let P be the maximal ideal of R and let Q be the maximal ideal of A. Since A is a finitely generated R-module, $PA \neq A$ by Nakayama's lemma, so $PA \subset Q$. Since A is finitely generated over R, A/PA is a finite-dimensional R/P vector space, and is thus Artinian. It follows that $Q^n \subset PA$, so $\operatorname{depth}(Q, A) = \operatorname{depth}(PA, A)$.

If $x_1, \ldots, x_n$ is a maximal A-sequence in P, then P is contained in an associated prime (in R) of $A/(x_1, \ldots, x_n)A$, and thus P annihilates some nonzero element $y \in A/(x_1, \ldots, x_n)A$. It follows that PA annihilates y, so $x_1, \ldots, x_n$ is a maximal A-sequence in PA. This shows $\operatorname{depth}(PA, A) = \operatorname{depth}(P, A)$.

By the Auslander-Buchsbaum formula, $\operatorname{depth}(P, A) = \operatorname{depth} R - \operatorname{pd}_R A = \dim R - \operatorname{pd}_R A$. Putting these things together we see that $\operatorname{depth}(Q, A) = \dim A$ iff $\operatorname{pd}_R A = \operatorname{codim}_R A$. □

19.4 Stably Free Modules and Factoriality of Regular Local Rings

Another success of "representation theory" is the theorem of Auslander and Buchsbaum that regular local rings are factorial. It turns out that this is a consequence of an odd remark.

Proposition 19.16 (Serre). *Any projective module with a finite free resolution is stably free—that is, it becomes free after adding a free module.*

Proof. If

$$\mathcal{F}:\ 0 \to F_n \to F_{n-1} \to \cdots \to F_0 \to P \to 0$$

is a free resolution of the projective module P, then the epimorphism $F_0 \to P$ must split. Thus the kernel, which is the image of F_1, is also projective, and continuing this way, we see that each F_i is the direct sum of its image in F_{i-1} and the image of F_{i+1}. Consequently, we have

$$P \oplus F_1 \oplus F_3 \oplus \cdots \cong F_0 \oplus F_2 \oplus \ldots,$$

that is, $P\oplus$ free module $\cong$ free module.

With Corollary 19.8, Proposition 19.16 shows that if k is a field, then projective modules over a polynomial ring $k[x_1, \ldots, x_n]$ are stably free. In

fact, they *are* free. This is the algebraic analogue of saying that vector bundles on an affine space are trivial, an obvious fact because affine space is contractible. Serre [1955], observing this analogy, asked whether the algebraic fact would hold too, and he proved Proposition 19.16 as a first bit of evidence that it might. The problem had quite an active history—interest in it contributed, for example, to the early development of algebraic K-theory—and was finally laid to rest by Quillen and Suslin independently in 1976—see Lam [1978] for a beautiful exposition.

Lest the reader imagine that every stably free module is free, we give a classic example:

Example 19.17. Let

$$R = \mathbf{R}[x_1, \ldots, x_n]/(1 - \sum x_i^2),$$

and let T be the R-module with free presentation

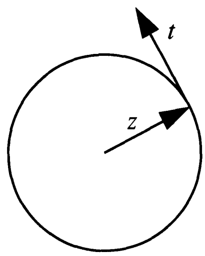
$$0 \to R \xrightarrow{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}} R^n \to T \to 0.$$

It is quite easy to see that T is stably free: Indeed, since $\sum x_i^2 = 1$ in R, the map $R^n \to R$ given by the matrix $(x_1, \ldots, x_n)$ splits the map $R \to R^n$, so $T \oplus R \cong R^n$. Under this identification, T consists of the vectors

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \in R^n$$

such that $\sum x_i t_i = 0$.

It is a remarkable fact, still lacking a simple algebraic proof, that T is actually free if and only if $n = 1, 2, 4,$ or 8. To see at least part of this, note first that R is the affine coordinate ring of the real $(n-1)$-sphere. T is in fact the module of "polynomial sections of the tangent bundle": Let $t \in T \subset R^n$. If z is a point of the $(n-1)$-sphere, thought of as a unit vector in $\mathbf{R}^n$, so that $P_z := (x_1 - z_1, \ldots, x_n - z_n)$ is a maximal ideal of R, then the image of t in $R^n/P_z R^n = \mathbf{R}^n$ is a vector with coordinates t_i satisfying $\sum z_i t_i = 0$, that is, a vector orthogonal to z. Such vectors may be identified with tangent vectors to the sphere at the point z,



so that t corresponds to a tangent vector field to the sphere. Since the t_i are polynomial functions of the coordinates of z, t is a polynomial vector field, and following the same construction backward, it is clear that every polynomial vector field corresponds to a section of T.

Now if T is free, then we can choose a basis, which will consist of $n-1$ elements of T whose images in T/P_zT form a basis for every z. Thinking of these as vector fields, we see that they yield a trivialization of the tangent bundle to the $(n-1)$-sphere. However, the tangent bundle is only trivial—even topologically—iff $n = 1, 2, 4,$ or 8. Essentially it is trivial in these cases because then the sphere is the group of elements of norm 1 in the reals, complexes, quaternions, and Cayley numbers, respectively; the group law makes possible a "parallel transport" of a basis of the tangent space at any one point. Since the multiplication laws in these algebras are given by polynomials, the trivializing vector fields may be taken to be sections of T. The proof that these are the only cases where the tangent bundle is trivial is very deep.

One can go a little further in this direction: For example, you can't even "comb down hair" on a two-sphere, so even a rank 1 free direct summand in case $n = 3$ does not exist. For an account of the topological results we have used, see Husemoller [1975].

In contrast to this difficult situation, it is easy to see that a stably free ideal is always free. We shall use this fact in the proof that regular local rings are factorial.

Lemma 19.18. *If M is a module such that $M \oplus R^{n-1} \cong R^n$, then $M \cong R$.*

Proof. If $M \oplus R^{n-1} \cong R^n$, then by Proposition A2.2c,

$$\begin{aligned} R &\cong \wedge^n R^n \\ &\cong \wedge^n (M \oplus R^{n-1}) \\ &\cong \sum_{i+j=n} \wedge^i M \otimes \wedge^j R^{n-1} \\ &\cong M \oplus \wedge^2 M \otimes \wedge^{n-2} R^{n-1} \oplus \cdots . \end{aligned}$$

If now P is any maximal ideal of R, then M_P is free of rank 1, so $\wedge^i M_P = 0$ for all $i \geq 2$. It follows that $\wedge^i M = 0$ for all $i \geq 2$, and we are done. □

See Exercises 19.4 and 19.6 for a sharpening and application of this idea.

We are now ready for the factoriality result.

Theorem 19.19. *Every regular local ring is factorial.*

A more refined version (MacCrae [1965]; see also Buchsbaum and Eisenbud [1974]) says that a ring is factorial iff every two-generated ideal has finite projective dimension.

Proof. Let R be a regular local ring, and let x be an element of a minimal set of generators of the maximal ideal of R. Since $R/(x)$ is again a regular local ring, x is prime by Corollary 10.14. We shall use the following:

Lemma 19.20 (Nagata). *If x is a prime element of an integral domain R, and if a prime Q not containing x becomes principal in $R[x^{-1}]$, then Q is principal. In particular, if $R[x^{-1}]$ is factorial, then R is factorial.*

Proof. Choose an element $q \in Q$ such that $qR[x^{-1}] = QR[x^{-1}]$, and the ideal $(q) \subset R$ is maximal with this property—in particular, we may assume that $q \notin (x)$. It suffices to show that if $xy \in (q)$ then $y \in (q)$—that is, that x is a nonzerodivisor $\text{mod}(q)$.

Write $xy = rq$. Since (x) is prime, $q \notin (x)$, we must have $r = sx$ for some s. Dividing both sides by x we get $y = sq$, as required.

To deduce the final statement we apply Corollary 10.6, noting that the codimension-1 primes of R are precisely (x) and the codimension-1 primes Q not containing x. □

We assume again that R is regular, and prove Theorem 19.19 by induction on dim R. By Lemma 19.20 it is enough to prove that $R[x^{-1}]$ is factorial, that is, that every codimension-1 prime Q of $R[x^{-1}]$ is principal. If P is a maximal ideal of $R[x^{-1}]$, then $R[x^{-1}]_P$ is also a localization of R, so by Corollary 19.14, it is a regular local ring. Its dimension is less than that of R, so by induction we may assume that it is factorial. Thus Q_P is principal, and thus a free R_P-module for every P. Theorem 19.2 shows that Q is projective as an $R[x^{-1}]$-module; we must prove that it is actually free.

Of course Q is the localization of an ideal Q' of R, so Q has a finite free resolution over $R[x^{-1}]$, obtained by localizing a free resolution of Q'. Thus by Corollary 19.16, Q is stably free, and by Lemma 19.18, Q is free. □

19.5 Exercises

Throughout these exercises we assume that all rings considered are Noetherian.

Regular Rings

Exercise 19.1: Show that the completion of a regular local ring is again regular.

Exercise 19.2: Suppose that a local ring R is a complete intersection in the sense that $R = S/I$, where S is a regular local ring and I is generated by a regular sequence. Show that any localization of R has the same property.

Exercise 19.3:* Imitate the proofs of Corollaries 19.7 and 19.8 to prove that a Noetherian ring R is regular iff the polynomial ring $R[x]$ is regular.

Modules over a Dedekind Domain

Exercise 19.4:* Let R be a domain. If $I_1, \dots, I_n$ and $J_1, \dots, J_m$ are nonzero ideals with

$$I_1 \oplus \cdots \oplus I_n \cong J_1 \oplus \cdots \oplus J_m,$$

then $n = m$ and $I_1 \cdot \cdots \cdot I_n \cong J_1 \cdot \cdots \cdot J_n$. (Something like the assumption that R is a domain is necessary to avoid cases like the one where R has a nontrivial idempotent e, and $I_1 = Re$, $I_2 = R(1-e)$, so $R = I_1 \oplus I_2$. It is enough to assume that each I_j contains a nonzerodivisor. However, if we assume $m = n$, then the statement is true—but much more subtle—without the assumption about nonzerodivisors. See Heitmann and Wiegand [1991].)

Exercise 19.5: Let R be a Dedekind domain. If $I_1, \dots, I_t$ are nonzero ideals of R, show that $\oplus_{j=1}^t I_j \cong R^{t-1} \oplus \Pi_{j=1}^t I_j$. Thus if $I_1, \dots, I_n$ and $J_1, \dots, J_m$ are nonzero ideals, then

$$I_1 \oplus \cdots \oplus I_n \cong J_1 \oplus \cdots \oplus J_m,$$

iff $n = m$ and $I_1 \cdot \cdots \cdot I_n \cong J_1 \cdot \cdots \cdot J_n$.

The problem reduces immediately to the case $n = 2$. The main point: Given two invertible modules I_1, I_2, use localization to show that they can be embedded in R in such a way that $I_1 + I_2 = R$. Form an exact sequence $0 \to J \to I_1 \oplus I_2 \to R \to 0$, and use Exercise 19.4 to show that $J \cong I_1 I_2$.

Exercise 19.6:* Suppose that R is a Dedekind domain.

a. Let P be a torsion-free R-module. Show that P is projective. Show that P is isomorphic to a unique module of the form $R^t \oplus I$, where I is an ideal.

b. Let M be any finitely generated R-module, and let M_{tors} be the torsion submodule of M —that is, the set of elements of M annihilated by some nonzero element of R. Show that $M \cong M/M_{\text{tors}} \oplus M_{\text{tors}}$.

c. Let M be a torsion R-module (that is, $M = M_{\text{tors}}$). Show that M may be written uniquely in the form

$$M = R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_m,$$

where $0 \neq I_1 \subset \cdots \subset I_m \subsetneqq R$ is an ascending sequence of ideals of R.

The Auslander-Buchsbaum Formula

Exercise 19.7: An elementary special case of the Auslander-Buchsbaum formula: Let (R, P) be a 0-dimensional local ring. Prove directly that no submodule of PR^n can be free. Conclude that any module of finite projective dimension over R is free.

Exercise 19.8 (Auslander-Buchsbaum formula in the graded case): Let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded ring, finitely generated as an algebra over a field R_0. Let $P = R_1 \oplus R_2 \oplus \cdots$ be the homogeneous maximal ideal. If M is a finitely generated graded R-module of finite projective dimension, then $\operatorname{pd} M = \operatorname{depth}(P, R) - \operatorname{depth}(P, M)$.

Projective Dimension and Cohen-Macaulay Rings

Exercise 19.9: An ideal $I \subset R$ is said to be **perfect** if the inequality of Corollary 18.5 is an equality for $M = R/I$; that is, if depth $(I, R) = \operatorname{pd}_R R/I$. Ideals generated by regular sequences, among others, are perfect. Show that if R is Cohen-Macaulay, and I is perfect, then R/I is Cohen-Macaulay.

Exercise 19.10: Use Proposition 3.12, Corollary 19.15, and Exercise 19.8 to show that if $R = k[x_0, \ldots, x_n]/I$ is a graded ring, then R is Cohen-Macaulay iff R_P is Cohen-Macaulay, where $P = (x_0, \ldots, x_n)$.

Exercise 19.11: If R is Cohen-Macaulay, then R has a module of finite length and finite projective dimension. (The converse is a well-known conjecture of Bass, known to be true for rings containing a field.)

Exercise 19.12:* Theorem 19.2 is false if we do not assume that M is finitely generated, even over rings as simple as $k[x]$ or $\mathbf{Z}$: Let M be the $\mathbf{Z}$-module consisting of all rational numbers with square-free denominator. Show that M becomes free when localized at any prime ideal of $\mathbf{Z}$. Show that M is not projective.

Exercise 19.13: For any nonzero module M, and any prime $P \in \operatorname{Ass} M$, show that $\operatorname{pd} M \geq \operatorname{depth} P$.

Hilbert Function and Grothendieck Group

Let $S = k[x_1, \ldots, x_r]$, where x_i is an indeterminate of degree d_i. Set $q(t) = \Pi_{i=1}^r (1 - t^{d_i})$. Recall from Exercises 10.11–10.13 that the **Hilbert series** of M is the formal power series in one variable t given by $h_M(t) := \sum_{n \geq 0} H_M(n) t^n$.

Exercise 19.14:

a. Show that the Hilbert series of S is given by $h_S(t) = 1/q(t)$, and thus that the Hilbert series of $S(-a)$ is $h_{S(-a)}(t) = t^a/\Pi(1-t^{d_i})$.

b. Let M be a finitely generated graded S-module. From Corollary 19.7 we know that M has free resolution of the form

$$0 \to F_m \to \cdots \to F_1 \to F_0 \to M \to 0,$$

with $F_i = \oplus S(-a_{ij})$. Let $p(t) = \sum_{i,j}(-1)^i t^{a_{ij}}$. Show that $h_M(t) = p(t)/q(t)$.

Exercise 19.15 (The Hilbert series is universal on graded modules): Let $S = k[x_1, \ldots, x_r]$, where k is a field and (for simplicity) all the indeterminates have degree 1. Let $\mathcal{C}$ be the category of finitely generated graded S-modules. The Hilbert series is an "additive function on $\mathcal{C}$" with values in the (additive group of) formal power series $\mathbf{Z}[[t]]$ in the sense that for each module $M \in \mathcal{C}$ we have the power series $h_M(t) = \sum_{d=0}^{\infty} \dim_k(M_d) \in \mathbf{Z}[[t]]$, and for each short exact sequence $0 \to M' \to M \to M'' \to 0$ in $\mathcal{C}$ we have $h_{M''} - h_M + h_{M'} = 0$ in $\mathbf{Z}[[t]]$. In fact, it is the **universal additive function**, in a certain sense. We prove this now:

We define the **Grothendieck group** $G_0(\mathcal{C})$ to be the additive group with a generator $[M]$ for each graded module M in $\mathcal{C}$, and a relation $[M''] - [M] + [M']$ for each short exact sequence $0 \to M' \to M \to M'' \to 0$ in $\mathcal{C}$.

Show that the map $M \mapsto [M]$ is the universal additive function on $\mathcal{C}$ in the sense that given any other additive function h with values in a group A, there is a unique group homomorphism $h' : G_0(\mathcal{C}) \to A$ such that $h(M) = h'([M])$. In particular, the Hilbert series induces a function on $G_0(\mathcal{C})$. Show that this function is a monomorphism on $G_0(\mathcal{C})$, and maps $G_0(\mathcal{C})$ isomorphically to $(1-t)^{-r}(\mathbf{Z}[t, t^{-1}]) \subset \mathbf{Z}[[t]]$, using the following steps:

a. Define $K_0(\mathcal{C})$ to be the additive group with a generator $[F]$ for each graded free module F, and a relation $[F''] - [F] + [F']$ for each short exact sequence $0 \to F' \to F \to F'' \to 0$ in $\mathcal{C}$. (Since every short exact sequence of free modules splits, we could instead have taken a relation $[F''] - [F] + [F']$ whenever $F \cong F' \oplus F''$ as graded modules.) Show that regarding a free module as a module gives a map of groups $K_0(\mathcal{C}) \to G_0(\mathcal{C})$. Because of the existence of finite free resolutions of a graded module by graded free modules, this map has an inverse. Construct it.

b. Use part a to show that $G_0(\mathcal{C}) = K_0(\mathcal{C})$ is the free group on the classes of modules $S(d)$ for $d \in \mathbf{Z}$. Now use Ex. 19.14a.

Exercise 19.16 (The Hilbert polynomial is universal on sheaves): The Hilbert polynomial has a universal description too: The

map taking each graded module M to its Hilbert polynomial is the universal additive function on modules that is zero on modules of finite length. Prove this as follows:

a. With notation as in Exercise 19.15, let $\bar{G}_0(\mathcal{C}) = G_0(\mathcal{C})/\mathcal{F}$, where $\mathcal{F}$ is the subgroup generated by the classes of modules of finite length. (If you like that language, you might think of $\bar{G}_0(\mathcal{C})$ as the Grothendieck group of coherent sheaves on $\mathbf{P}^{r-1}$, and $P_M(n)$ as the Euler characteristic of the sheaf associated to M.) Show that the map of sets $M \mapsto P_M(n)$ taking each graded module to its Hilbert polynomial is a group homomorphism from $\bar{G}_0(\mathcal{C})$ to the additive group of $\mathbf{Q}[n]$.

b. Show that the Koszul complex of $x_1, \ldots, x_r$ leads to a relation

$$\sum_{n=0}^{r} (-1)^n \binom{r}{n} [S(-n)] = 0 \quad \text{in } \bar{G}_0(\mathcal{C}).$$

c. Show that $\bar{G}_0(\mathcal{C})$ is generated by the classes $[S(-a)]$ for $a = 0, \ldots, r-1$.

d. Show from Exercise 19.14a that the Hilbert series of M may be written in the form $h_M(t) = f_M(t)/(1-t)^r + g_M(t)$, where $f_M(t)$ is a polynomial of degree $< r$, and $g_M(t)$ is a polynomial, uniquely determined by this relation. Show that $f_{S(-a)} = t^a$ for $0 \le a \le r-1$. Deduce that $\bar{G}_0(\mathcal{C})$ is the free abelian group on the classes $[S(-a)]$ for $a = 0, \ldots, r-1$.

e. Now prove that the map $[M] \mapsto P_M(n) \in \mathbf{Q}[n]$ induces an isomorphism from $\bar{G}_0(\mathcal{C})$ onto its image.

Exercise 19.17: In terms of the basis for $\bar{G}_0(\mathcal{C})$ constructed in Exercise 19.16, compute the class $[S/(x_1, \ldots, x_t)]$ for each $t < r$. Do these classes form a basis for $\bar{G}_0(\mathcal{C})$?

The Chern Polynomial

Exercise 19.18:* Let S be the graded polynomial ring $k[x_1, \ldots, x_{r+1}]$, where each x_i has degree 1, the homogeneous coordinate ring of $\mathbf{P}^r$ over the field k. Suppose the degrees of the x_i are all equal to 1, so that the graded ring S corresponds to projective space $\mathbf{P}^r_k$. Let M be a finitely generated graded S-module. In geometry the coefficients of the Hilbert polynomial are usually coded in a different way, as **Chern classes** (or **Chern numbers**). These may be defined as the coefficients of the **Chern polynomial** $c_t(M)$ of M (or actually of the sheaf associated to M on projective space), which we shall now describe.

The Chern polynomial $c_t(M)$ is an element of $\mathbf{Z}[t]/(t^{r+1})$ defined by the following two properties. (See Fulton [1984, Chapter 3]; Chern classes are

defined there initially only for vector bundles, but for nonsingular varieties such as $\mathbf{P}^{r-1}$ one may extend the definition to all sheaves by using free resolutions.)

i. The Chern polynomial of $S(-a)$ is defined to be $c_t(S(-a)) = 1 - at$.

ii. The Chern polynomial is multiplicative in exact sequences, in the sense that if $0 \to M' \to M \to M'' \to 0$ is a short exact sequence, then $c_t(M) = c_t(M')c_t(M'')$.

Prove that there exists one and only one polynomial satisfying conditions i and ii.

The Hirzebruch-Riemann-Roch formula for sheaves on projective space (see Fulton [1984, Chapter 15]) provides a way of computing the Chern and Hilbert polynomials from one another. It seems plausible that there is a simple combinatorial expression of this relationship, but I do not know of one.

20

Free Resolutions and Fitting Invariants

We shall treat five related areas in this chapter:

1. Uniqueness of free resolutions
2. Fitting ideals of modules
3. What makes a complex exact?
4. The Hilbert-Burch structure theorem for perfect ideals of codimension 2
5. Castelnuovo-Mumford regularity

Many invariants in algebraic geometry and commutative algebra—from intersection numbers of varieties to the cohomology of sheaves to the depth and dimension of a module—may be defined in terms of free resolutions. Many of these invariants are actually invariants of the homology of complexes derived from free resolutions. Some others seem accessible only through free resolutions themselves.

In this chapter we introduce some topics related to finite free resolutions. We begin with the uniqueness of the minimal free resolution of a module over a local ring. Then we study Fitting invariants, which generalize the structure theory of modules over a principal ideal domain and, in general, give a way of expressing certain features of a module in terms of ideals. In the third section we use the Fitting ideals of syzygy modules to characterize resolutions among all finite complexes of free modules. In the fourth section we apply this information to give a structure theorem due to Hilbert and

Burch for certain resolutions of length 2 that has been the beginning for a great deal of recent work. Finally, in the last section, we explain a few facts about the "regularity" of graded modules, an idea due to Mumford that is a better-behaved version of the degree of the generators of a module.

The results here involving minimal resolutions all work equally well in the local case and in the case of a positively graded algebra over a field. As pointed out by Goto and Watanabe [1978], there is a natural common generalization of these two cases, "generalized local rings," described in Exercise 20.1. To avoid repeating everything, we stick with the local case in the text except in the discussion of regularity, which requires a grading, and in a few other applications. The reader may check that everything in the first four sections goes through in the generalized local case.

20.1 The Uniqueness of Free Resolutions

We have described how to produce a free resolution of any module by choosing generators of the module, then generators for the kernel of the resulting map from a free module, and so on. Of course, the resolution produced may not be finite, even if the module has finite projective dimension, and has no obvious uniqueness. Schanuel's lemma (Exercise A3.13) shows that the modules in a projective resolution do have a sort of stable uniqueness. But this is a weak result: In the local or graded cases the modules are in any case free. Exercise 20.16 contains an example illustrating the nonuniqueness that can occur. However we shall now show that in the local and graded cases the *minimal* free resolution is unique, and any free resolution is the direct sum of the minimal free resolution and a free resolution of the module 0.

The opposite of a minimal complex is a **trivial complex**, by which we mean a direct sum of complexes of the form

$$0 \to R \xrightarrow{1} R \to 0 \to 0 \cdots .$$

Thus, for example, the complex

$$0 \to R \xrightarrow{\binom{1}{0}} R^2 \xrightarrow{(0,1)} R \to 0 \cong \begin{array}{ccccccc} 0 & \to & R & \xrightarrow{1} & R & \to & 0 \\ & & \oplus & & \oplus & & \oplus \\ & & 0 & \to & R & \underset{1}{\to} & R & \to & 0 \end{array}$$

is a trivial complex. Trivial complexes have no homology, so if

$$\mathcal{G} : \; \cdots \to G_n \to \cdots \to G_0 \to 0$$

is a trivial complex and $\mathcal{F}$ is a resolution of a module M, then $\mathcal{F} \oplus \mathcal{G}$ is again a resolution of M. This is the simplest reason for the nonuniqueness of free resolutions. Over a local ring we shall see that it is the only reason.

Lemma 20.1. *If*

$$\mathcal{H}: \quad \cdots \to H_n \to \cdots \to H_1 \to H_0 \to 0$$

is a complex of free modules with trivial homology (that is, a free resolution of 0) over a local ring, then $\mathcal{H}$ *is a trivial complex.*

Proof. Since H_0 is free, the epimorphism $H_1 \to H_0$ splits, and we can write $H_1 = H_0 \oplus H_1'$, with

$$H_1' = \ker(H_1 \to H_0) = \operatorname{im}(H_2 \to H_1),$$

the map $H_1 \to H_0$ being the projection onto the first factor. Thus $\mathcal{H}$ is the direct sum of the complexes

$$\mathcal{H}_1 : 0 \to H_0 \xrightarrow{1} H_0 \to 0$$

and

$$\mathcal{H}': \quad \cdots \to H_n \to \cdots \to H_2 \to H_1' \to 0.$$

The complex $\mathcal{H}_1$ is a trivial complex, and the complex $\mathcal{H}'$ is at least a complex of projective modules with trivial homology with nonzero terms only in degrees greater than 0. Every projective module over a local ring is free (by Nakayama's Lemma, in the form of Exercise 4.11a) so $\mathcal{H}'$ is actually a free complex.

By the same argument we may write $\mathcal{H}' = \mathcal{H}_2 \oplus \mathcal{H}''$ and thus $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}''$ where $\mathcal{H}_1$ and $\mathcal{H}_2$ are trivial complexes and $\mathcal{H}''$ is a free complex with trivial homology, and nonzero terms only in degrees greater than 1. Continuing indefinitely we obtain a decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots$ as a direct sum of trivial complexes.

We are now ready for the uniqueness result.

Theorem 20.2 (Uniqueness of minimal free resolutions). *Let R be a local ring, and let M be a finitely generated R-module. If $\mathcal{F}$ is a minimal free resolution of M, then any free resolution of M is isomorphic to the direct sum of $\mathcal{F}$ and a trivial complex. In particular, there is up to isomorphism only one minimal free resolution of M.*

The first requirement for the proof of this result is to find a map between $\mathcal{F}$ and another given resolution of M. Such a map is induced from the identity map $1\colon M \to M$ as in the following easy lemma. Since it is a special case of Proposition A3.13 we shall omit the proof.

Lemma 20.3 (Maps from projective to acyclic complexes). *Let*

$$\mathcal{F}: \quad \cdots \to F_n \to \cdots \to F_0 \to M \to 0$$

be any complex of modules with each F_i projective, and let

$$\mathcal{G}: \quad \cdots \to G_n \to \cdots \to G_0 \to N \to 0$$

be a complex of modules without homology (for example, a free resolution of N).

If $\varphi : M \to N$ is any map, then there is a "lifting" of φ to a map of complexes

$$\bar{\varphi} : \mathcal{F} \to \mathcal{G}, \text{ inducing } \varphi : M \to N.$$

If $\tilde{\varphi}$ is another such lift, then $\bar{\varphi}$ and $\tilde{\varphi}$ are ***homotopic*** *in the sense that there exists a collection of maps $s_i : F_i \to G_{i+1}$ such that if we write $d_i : F_i \to F_{i-1}$ and $\delta_i : G_i \to G_{i-1}$ for the differentials of the complexes $\mathcal{F}$ and $\mathcal{G}$, respectively, then*

$$\bar{\varphi}_i - \tilde{\varphi}_i = \delta_{i+1} s_i + s_{i-1} d_i : F_i \to G_i. \qquad \square$$

Proof of Theorem 20.2. Let $\mathcal{G}$ be another free resolution of M. By Lemma 20.3 there are comparison maps $\alpha : \mathcal{F} \to \mathcal{G}, \beta : \mathcal{G} \to \mathcal{F}$ lifting the identity map on M. Further, $\beta\alpha$ is homotopic to 1: That is there exist s_i satisfying the relations of Lemma 20.3. Thus $1 - \beta_i\alpha_i : F_i \to F_i$ is a sum of maps factoring through the differentials of $\mathcal{F}$, and since $\mathcal{F}$ is minimal, this means that

$$(1 - \beta_i\alpha_i)(F_i) \subset PF_i,$$

where P is the maximal ideal of R. It follows that det $\beta_i\alpha_i \equiv 1 (\text{mod } P)$, so $\beta\alpha$ is an automorphism of $\mathcal{F}$. Replacing β with the composition of β and the inverse of this automorphism, we may assume that $\beta\alpha = 1$ —that is, α is an inclusion and the new β is a splitting of α.

Let $\mathcal{H} :=$ coker α. Since each α_i is a split inclusion, the modules of $\mathcal{H}$ are all free. Since the splitting of β is a map of complexes, $\mathcal{G} \cong \mathcal{F} \oplus \mathcal{H}$. Thus the homology of $\mathcal{G}$ is the direct sum of the homologies of $\mathcal{F}$ and $\mathcal{H}$. Since the inclusion $\alpha : \mathcal{F} \to \mathcal{G}$ induces an isomorphism on homology, the complex $\mathcal{H}$ has no homology at all. By Lemma 20.1, $\mathcal{H}$ is a trivial complex, and $\mathcal{G}$ is the direct sum of $\mathcal{F}$ with a trivial complex, as claimed. $\square$

20.2 Fitting Ideals

As a consequence of the uniqueness theorem just proved, we can make many invariants of a module out of a free (or even projective) resolution of the module. We begin by studying some that come from the first step of a free resolution, a "free presentation." Since this may be applied to the syzygy modules of a given module too, it will serve us later in the study of the whole free resolution. We shall assume throughout that our modules are finitely generated, and leave the adaptation to the infinite case in the hands of the interested reader.

Definition. *If $\varphi : F \to G$ is a map of free modules, then $I_j\varphi$ is the image of the map*

$$\wedge^j F \otimes \wedge^j G^* \to R$$

induced by $\wedge^j\varphi : \wedge^j F \to \wedge^j G$. *If we choose bases for* F *and* G, *then* φ *may be represented by a matrix, and we see that* $I_j\varphi$ *is generated by the minors (that is, determinants of submatrices) of size* j *of that matrix. We make the convention that the determinant of the* 0×0 *matrix (like a product with no factors) is 1. In particular,* $I_0\varphi = R$, *and more generally, we set* $I_j\varphi = R$ *if* $j \leq 0$.

It is not really necessary to suppose that F is free to make the preceding construction—we can replace F by any free module mapping onto F without changing $I_j\varphi$.

These ideals of minors turn out to define invariants of a module that generalize the usual invariants for finitely generated abelian groups, a fact that was first observed by Fitting [1936].

Corollary–Definition 20.4 (Fitting's Lemma). *Let* M *be a finitely generated module over a ring* R, *and let* $\varphi : F \to G \to M \to 0$ *and* $\varphi' : F' \to G' \to M \to 0$ *be two presentations, with* G *and* G' *finitely generated free modules of ranks* r *and* r'. *For each number* i *with* $0 \leq i < \infty$, *we have* $I_{r-i}(\varphi) = I_{r'-i}(\varphi')$, *and we define the* i*th Fitting invariant of* M *to be the ideal*

$$\mathrm{Fitt}_i(M) = I_{r-i}\varphi \subset R.$$

Proof. We leave to the reader the immediate reduction to the case where F and F' are finitely generated, the only case with which we shall be concerned—see Exercise 20.5.

Two ideals are equal iff they become equal in every localization of R, so we may harmlessly assume that R is local, and we must show that the Fitting ideals coming from a given free presentation of M are the same as the ones coming from the minimal presentation. If φ is the map giving the minimal presentation, then by Theorem 20.2 any other presentation map ψ may be put in the form

$$\psi = \left(\begin{array}{c|c|c} \varphi & 0 & 0 \\ \hline 0 & \mathbf{1} & 0 \end{array}\right),$$

where $\mathbf{1}$ is a $p \times p$ identity matrix. We must show that $I_j\varphi = I_{j+p}\psi$. Any nonzero minor m of ψ of size $j+p$ is made by taking, for some j', p' with $j' + p' = j + p$, a $j' \times j'$ minor m' of φ and a $p' \times p'$ minor of $\mathbf{1}$, and multiplying them. Since we must have $p' \leq p$, it follows that $j' \geq j$, and $m = m'$. Thus $I_{j+p}\psi = \sum_{j \leq j' \leq j+p} I_{j'}\varphi$. Since $I_{j'}\varphi \subset I_j\varphi$ for all $j' \geq j$, we are done.

The preceding proof—and indeed the whole theory of Fitting invariants—works just as well for projective presentations $F \to G \to M \to 0$ as for free ones as long as G has **constant rank** (that is, the same rank at every localization). This is true of any projective unless R contains nontrivial idempotents; see Exercise 20.12.

The Fitting ideals are functorial:

Corollary 20.5. *The formation of Fitting ideals commutes with "base change"; that is, for any map of rings $R \to S$,*

$$\operatorname{Fitt}_j(M \otimes_R S) = (\operatorname{Fitt}_j M)S.$$

Proof. By the right-exactness of tensor product, the tensor product of S with a free presentation of M is a free presentation of $S \otimes_R M$.

The significance of the Fitting ideals is perhaps best described by saying that the Fitting ideal $\operatorname{Fitt}_j M$ is the **obstruction to generating M with j elements** in the following sense.

Proposition 20.6. *If (R, P) is local, then M can be generated by j elements iff $\operatorname{Fitt}_j M = R$. In general, the closed subset of $\operatorname{Spec} R$ defined by $\operatorname{Fitt}_j M$ is the set of primes Q such that M_Q cannot be generated by j elements.*

Proof. By Corollary 20.5 the formation of the Fitting ideals commutes with localiztion, so we may suppose that R is local, and we need only show that M can be generated by j elements iff $\operatorname{Fitt}_j M = R$.

Let $\varphi : F \to G$ be a minimal presentation of M—that is coker $\varphi = M$ and φ is represented by a matrix with its entries in the maximal ideal P of R. Computing $\operatorname{Fitt}_j M$ from φ we see that $\operatorname{Fitt}_j M \subset P$ iff $j < \operatorname{rank} G$, the minimal number of generators of M.

The property given in Proposition 20.6 characterizes the Fitting ideals up to radical only. Because the generic determinantal ideals are prime however, the Fitting ideals themselves may be characterized as the collection of the largest ideals satisfying both Corollary 20.5 and Proposition 20.6. We shall not require this point, so we shall not pursue it further.

In the case of the zeroeth Fitting ideal, Proposition 20.6 shows that $\operatorname{Fitt}_0 M$ has the same radical as the annihilator of M. One can be more specific:

Proposition 20.7. *For any R-module M:*

a. $\operatorname{Fitt}_0 M \subset \operatorname{ann} M$.

b. For every $j > 0$ we have $(\operatorname{ann} M) \operatorname{Fitt}_j(M) \subset \operatorname{Fitt}_{j-1}(M)$. *If M can be generated by n elements, then* $(\operatorname{ann} M)^n \subset \operatorname{Fitt}_0 M$.

Proof. Suppose $M = \operatorname{coker} \varphi : G \to R^n$.

a. We have $\operatorname{Fitt}_0 M = I_n \varphi$. If φ' is an $n \times n$ submatrix of φ, then M is a homomorphic image of $M' := \operatorname{coker} \varphi'$, and thus $\operatorname{ann} M' \subset \operatorname{ann} M$. Thus it suffices to treat the case $G = R^n, \varphi = \varphi'$, and show that $\det \varphi \in \operatorname{ann} M$. Writing $\mathbf{1} : R^n \to R^n$ for the identity map, we may factor $(\det \varphi)\mathbf{1}$ as $\varphi \circ \psi$, where ψ is the matrix of cofactors of φ. The diagram

$$\begin{array}{ccccc} R^n & \xrightarrow{\varphi} & R^n & \longrightarrow & M \to 0 \\ & {\scriptstyle\psi}\swarrow & \downarrow {\scriptstyle(\det\varphi)\mathbf{1}} & & \downarrow {\scriptstyle(\det\varphi)|_M = 0} \\ R^n & \xrightarrow[\varphi]{} & R^n & \longrightarrow & M \to 0 \end{array}$$

commutes, so $(\det\varphi)M = 0$.

b. If $a \in \operatorname{ann} M$ and φ' is a $(j-1) \times (j-1)$ submatrix of φ with $j - 1 < n$, then we must show $a \cdot \det(\varphi') \in I_j(\varphi)$. We can make a new presentation matrix for M of the form $\varphi + aI : G \oplus R^n \to R^n$, where $aI : R^n \to R^n$ is multiplication by a. In terms of matrices, $\varphi + aI$ is represented by an $n \times (n+m)$ matrix of the form pictured in Figure 20.1, where $m = \operatorname{rank} G$. Let φ'' be the submatrix of $\varphi + aI$ obtained from φ' by adding a row not involved in φ' and adding the corresponding column from aI, as in Figure 20.1.

We have $\varphi'' = \varphi' \oplus a$, so $\det(\varphi'') = a \cdot \det(\varphi')$. By Fitting's lemma, Corollary 20.4, $\det\varphi'' \in I_j(\varphi + aI) = I_j(\varphi)$ as required.

Finally, note that $\operatorname{Fitt}_n(M) = R$, so from the preceding relation we successively deduce

$$\begin{aligned} &(\operatorname{ann} M)R \subset \operatorname{Fitt}_{n-1}(M), \\ &(\operatorname{ann} M)^2 \subset (\operatorname{ann} M)\operatorname{Fitt}_{n-1}(M) \subset \operatorname{Fitt}_{n-2}(M) \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ &(\operatorname{ann} M)^n \subset (\operatorname{ann} M)\operatorname{Fitt}_1(M) \subset \operatorname{Fitt}_0(M). \end{aligned}$$

□

The case $M = (R/I)^n$ shows that the exponent n in part b cannot be improved. However, the result can be considerably sharpened and extended. Some results in this direction are given in Exercises 20.6, 20.9, and 20.10, Buchsbaum and Eisenbud [1977], and Eisenbud and Green [1994].

If $\operatorname{ann} M = 0$, then Proposition 20.7 tells us only that $\operatorname{Fitt}_0 M = 0$. In general, if $M = \operatorname{coker}(\varphi : F \to G)$, then the most important of the Fitting ideals of M is the first of the $\operatorname{Fitt}_j M$ that is nonzero. We call this Fitting ideal simply $I(M)$ or, to make the notation more convenient in many cases, $I(\varphi)$. Thus $I(\varphi) = I_{\operatorname{rank}\varphi}\varphi$. Its significance for us is the following:

Proposition 20.8. *M is projective of constant* $\operatorname{rank} r$ *iff* $\operatorname{Fitt}_r(M) = R$ *and* $\operatorname{Fitt}_{r-1}(M) = 0$. *Thus M is projective of (some) constant rank iff $I(M) = R$.*

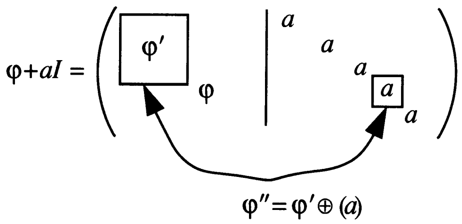


FIGURE 20.1.

Proof. Suppose that M is projective of constant rank r; that is, suppose that for all primes Q of R the module M_Q is free of rank r. Over R_Q the module M_Q has a free presentation $F \to G \to M_Q \to 0$ with $F = 0$ and $G = M_Q$, so $\mathrm{Fitt}_j M_Q = R_Q$ if $j \geq r$, while $\mathrm{Fitt}_j M_Q = 0$ otherwise. By Corollary 20.5 the Fitting ideals localize, so $\mathrm{Fitt}_j M = R$ if $j \geq r$, while $\mathrm{Fitt}_j M = 0$ otherwise. In particular, $I(M) = R$.

Conversely, suppose that $\mathrm{Fitt}_r M = R$ while $\mathrm{Fitt}_{r-1} M = 0$, and let Q be a prime. By Proposition 20.6, M_Q can be generated by r elements over R_Q. Let $M_Q = \mathrm{coker}(\varphi : G \to R_Q^r)$ be a free presentation over the localization R_Q of R. Since $0 = \mathrm{Fitt}_{r-1} M_Q = I_1(\varphi)$, we have $\varphi = 0$ and M_Q is free of rank r.

The construction of $I(\varphi) = I_{\mathrm{rank}\,\varphi}\varphi$ does not in general commute with localization since the rank of φ may decrease when we localize. However, this is the only problem, so that if $I(\varphi)$ contains a nonzerodivisor then $I(\varphi) \otimes R[S^{-1}] = I(\varphi \otimes R[S^{-1}])$ for any multiplicatively closed set S. We shall use this remark in the proof of Theorem 20.9.

20.3 What Makes a Complex Exact?

The next result shows that a finite free resolution can be distinguished from an arbitrary free complex simply by examining the ideals of minors of its maps.

Theorem 20.9. *Let R be a ring. A complex*

$$\mathcal{F} : 0 \to F_n \xrightarrow{\varphi_n} F_{n-1} \to \cdots \to F_1 \xrightarrow{\varphi_1} F_0$$

of free R modules is exact iff

(1) $$\mathrm{rank}\, F_k = \mathrm{rank}\, \varphi_k + \mathrm{rank}\, \varphi_{k+1}$$

and

(2) $$\mathrm{depth}\, I(\varphi_k) \geq k$$

for $k = 1, \ldots, n$.

Here we interpret the unit ideal as having infinite depth, so that if $I(\varphi_k) = R$, then condition (2) is automatically satisfied. We postpone the proof of Theorem 20.9 for a moment.

It is easy to make versions for projective modules and for the exactness of $\mathcal{F} \otimes M$, for any module M. For example, in the latter case the numbers that are important are $\mathrm{rank}(\varphi_k, M)$, the largest r such that $(I_r\varphi)M \neq 0$, and $\mathrm{depth}(I_r(\varphi), M)$. For these generalizations and others, see Buchsbaum and Eisenbud [1973] or Northcott [1976].

One might hope for a similar result for complexes with no 0 at the left-hand end, so that one could apply it to infinite resolutions as well, or simply to a pair of maps whose composition is zero, a complex of the form

$$F \xrightarrow{\varphi} G \xrightarrow{\psi} H.$$

Unfortunately, no such result is known. To get an idea how far this case may deviate from the conditions of the preceding theorems, even when the given complex forms part of a minimal free resolution over a local ring, consider the ring $R = k[\varepsilon]$ with $\varepsilon^2 = 0$, and the matrix

$$\varphi = \begin{pmatrix} \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{pmatrix}.$$

It is easy to check that over this ring the complex

$$\cdots \xrightarrow{\varphi} R^3 \xrightarrow{\varphi} R^3 \xrightarrow{\varphi} R^3 \xrightarrow{\varphi} \cdots$$

is exact although $\operatorname{rank} \varphi = 1$ in our sense, and $\operatorname{depth} I(\varphi) = 0$. See Eisenbud [1980] for more information on such resolutions.

Nonetheless, there is one elementary result of this kind that holds under a strong hypothesis and generalizes the case of vector spaces over a field. We shall use it in the proof of Theorem 20.9. For a slightly more general result, see Exercise 20.4.

Lemma 20.10. *A complex $F \xrightarrow{\varphi} G \xrightarrow{\psi} H$ of free R-modules with $I(\varphi) = I(\psi) = R$ is exact iff* $\operatorname{rank} \varphi + \operatorname{rank} \psi = \operatorname{rank} G$.

Proof. The most direct way to prove this lemma is to localize so that one may assume that φ and ψ have minors of size $\operatorname{rank} \varphi$ and $\operatorname{rank} \psi$ that are units, and then to reduce φ and ψ as far as possible to normal form. Since it is hard to give such an argument without a blackboard, we take a more formal approach:

We may harmlessly assume that R is local. In this circumstance, Proposition 20.8 shows that coker φ is free of rank equal to $\operatorname{rank} G - \operatorname{rank} \varphi$. The map ψ may be factored as

$$G \to \operatorname{coker} \varphi \xrightarrow{\psi'} H,$$

and the rank and ideals of minors of ψ' are the same as those of ψ. Thus it suffices to prove the result when $\varphi = 0$.

Write $-^*$ for $\operatorname{Hom}_R(-, R)$. By Proposition 20.8, coker (ψ^*) is free of rank equal to $\operatorname{rank}(G^*) - \operatorname{rank}(\psi^*) = \operatorname{rank} G - \operatorname{rank} \psi$. Dualizing the right-exact sequence

$$H^* \to G^* \to \operatorname{coker}\,(\psi^*) \to 0,$$

we get a left-exact sequence

$$0 \to \operatorname{coker}(\psi^*)^* \to G \to H.$$

The dual of a free module is 0 iff the module is 0, so $\psi : G \to H$ is a monomorphism iff $\operatorname{rank}\psi = \operatorname{rank} G$.

Proof of Theorem 20.9. We first show that if $\mathfrak{F}$ satisfies (1) and (2), then $\mathfrak{F}$ is exact. It is enough to prove this after localizing. Because (2) guarantees in particular that each $I(\varphi_k)$ contains a nonzerodivisor, the rank of each φ_k is preserved by localizing. Thus hypotheses (1) and (2) are preserved, and we may begin by supposing that R is local.

We may reduce to the case $n \leq d :=$ depth R. Since no proper ideal of R can have depth > depth R, condition (2) implies that $I(\varphi_k) = R$ for all $k >$ depth R. Thus Lemma 20.10 proves the exactness of the complex $\mathfrak{F}$ at F_k for $k > d$. Furthermore, coker φ_{d+1} is projective (and thus, since R is local, free) by Proposition 20.8, so we may simply replace F_d by coker φ_{d+1} and assume from now on that $n \leq d$.

By induction on the dimension of R, we may suppose that $\mathfrak{F}$ becomes exact when localized at any nonmaximal prime (admitting the case that R is zero-dimensional, so that there are no nonmaximal primes). Thus we may assume that for $i \geq 1$ the homology of $\mathfrak{F}$ at F_i is annihilated by a power of the maximal ideal, and in particular that it has depth 0. The following powerful result of Peskine and Szpiro [1973] completes the proof of the exactness of $\mathfrak{F}$ in the case $n \leq$ depth R (which is the depth of F_i for every i, since the F_i are free).

Lemma 20.11 (Acyclicity Lemma). *Suppose*

$$\mathfrak{F} : 0 \to F_n \xrightarrow{\varphi_n} F_{n-1} \to \cdots \to F_1 \xrightarrow{\varphi_1} F_0$$

is a complex of finitely generated (but not necessarily free) modules F_i over a local ring R such that depth $F_i \geq i$. *If $H_i\mathfrak{F} \neq 0$ for some $i > 0$, then for the largest such i we have* depth $H_i\mathfrak{F} \geq 1$.

Proof. Truncating the complex on the right if necessary, we may assume that $H_1\mathfrak{F}$ is the only nonvanishing homology. If $n = 1$, then $H_1\mathfrak{F} \subset F_1$ has depth ≥ 1 as required, so we may assume $n \geq 2$. Writing $K = \ker \varphi_1$ and $B_i = \operatorname{im} \varphi_i$, we have exact sequences

$$0 \to B_{k+1} \to F_k \to B_k \to 0, \quad \text{for } k \geq 2,$$

and

$$0 \to B_2 \to K \to H_1\mathfrak{F} \to 0.$$

Applying Corollary 18.6 iteratively to the first group of these sequences, starting from $k = n$ and depth F_n = depth $B_n \geq n$, we see that depth $B_k \geq k$ for all $n \geq k \geq 2$. On the other hand, depth $K \geq 1$ because K is a submodule of F_1. Using Corollary 18.6 again, we see that depth $H_1\mathfrak{F} \geq 1$ as required. ☐

Continuing with the proof of Theorem 20.9, we must show that if the complex $\mathfrak{F}$ is exact, then conditions 1) and 2) are satisfied. We shall localize

and use the Auslander-Buchsbaum formula. To use this technique we need to know that the formation of the $I(\varphi_k)$ commutes with localization. To this end we shall show that each $I(\varphi_k)$ contains a nonzerodivisor.

We may begin by inverting all the nonzerodivisors in R, and thus assume that R is a semilocal ring where all the maximal ideals are associated primes of R. In this situation we must show that $I(\varphi_k) = R$ for every k. Since we may truncate the resolution $\mathcal{F}$ on the right, it is enough to show that $I(\varphi_1) = R$.

Let P be a maximal ideal of R. Since P is an associated prime of 0 in R we have depth $R_P = 0$. Let $M = \operatorname{coker} \varphi_1$. By the Auslander-Buchsbaum formula, Theorem 19.9, the module M_P is free over R_P. It follows that the complex $\mathcal{F}_P$ is split exact, so the rank of M_P is given by

$$\operatorname{rank} M_P = \sum_{i\geq 0} (-1)^i \operatorname{rank} F_i,$$

independently of the prime P. This shows that M is projective of constant rank. From Proposition 20.8 we get the desired equality $I(\varphi_1) = R$.

We now prove (1). To this end we may begin by inverting all the nonzerodivisors of R without changing the ranks of the maps or the modules. By what we have just proven, we shall then have $I(\varphi_k) = R$ for each k, so that Lemma 20.10 gives the desired conclusion.

It remains to prove 2). Let P be a prime containing $I(\varphi_k)$ and associated to a maximal regular sequence in $I(\varphi_k)$, so that

$$\operatorname{depth}\,(P_P, R_P) = \operatorname{depth}\, I(\varphi_k).$$

Localizing $\mathcal{F}$ at P and using Proposition 20.8 on the module coker φ_k, we see that coker $(\varphi_k)_P$ is not free, so $\operatorname{pd}_{R_P} M_P \geq k$ by Proposition 20.2. Since M_P is a module of finite projective dimension, the Auslander-Buchsbaum formula shows that depth $(P_P, R_P) \geq k$. Our choice of P gives depth $I(\varphi_k) \geq k$, and we are done.

Here are some applications of Theorem 20.9:

Corollary 20.12. *Let $\mathcal{F}$ be a complex as in Theorem 20.9. If $\mathcal{F}$ is exact, then*

$$\operatorname{rad}\, I(\varphi_k) \subset \operatorname{rad}\, I(\varphi_{k+1}) \quad \textit{for all } k \geq 1.$$

If in addition coker φ_1 has a nonzero annihilator, then φ_1 has rank equal to the rank of F_0. In this case, writing $d = \operatorname{depth} I(\varphi_1)$, we have

$$\operatorname{rad}\, I(\varphi_1) = \cdots = \operatorname{rad}\, I(\varphi_d) \neq \operatorname{rad}\, I(\varphi_{d+1}).$$

Proof. By Theorem 20.9, the exactness of $\mathcal{F}$ implies that each of the ideals $I(\varphi_k)$ contains a nonzerodivisor, so φ_k has the same rank at every localization of R.

To prove the first statement it is enough to show that if a prime Q does not contain $I(\varphi_k)$ then it does not contain $I(\varphi_{k+1})$. But if Q does not

contain $I(\varphi_k)$ then $\text{coker}(\varphi_k)_Q$ is projective by Proposition 20.8. Thus in particular the sequence

$$0 \to \text{coker}(\varphi_{k+1})_Q \to (F_k)_Q \to \text{coker}(\varphi_k)_Q \to 0$$

splits, so $\text{coker}(\varphi_{k+1})_Q$ is projective and hence free. Using Proposition 20.8 again, we see that $I(\varphi_{k+1}) \not\subset Q$, as required.

Now suppose that $M = \text{coker}\ \varphi_1$ has nonzero annihilator. Let $a \in I(\varphi_1)$ be a nonzerodivisor. Proposition 20.8 shows that $M[a^{-1}]$ is projective of constant rank over $R[a^{-1}]$. Since $M[a^{-1}]$ has nonzero annihilator, $M[a^{-1}]$ must be 0. It follows that $\text{rank}\, \varphi_1 = \text{rank}\, F_0$.

To establish the equalities in the last statement of Corollary 20.12, set $I = \text{ann}(\text{coker}\, \varphi_1)$. Since $\text{rank}\, \varphi_1 = \text{rank}\, F_0$, we have $I(\varphi_1) = \text{Fitt}_0(\text{coker}\ \varphi_1)$, so by Proposition 20.7 the radical of I coincides with the radical of $I(\varphi_1)$.

Suppose that Q is a prime ideal not containing $I(\varphi_k)$ for some $k \leq \text{depth}\ I$. Since $I(\varphi_k) \not\subset Q$, we know from Proposition 20.8 that $(\text{coker}\ \varphi_k)_Q$ is free. We must show that $I \not\subset Q$, or equivalently that $(\text{coker}\ \varphi_1)_Q = 0$.

If we assume that $(\text{coker}\ \varphi_1)_Q \neq 0$, then by Corollary 18.5 we know that

$$\text{depth}\ I_Q \leq \text{pd}(\text{coker}\ \varphi_1)_Q.$$

Since $(\text{coker}\ \varphi_k)_Q$ is free, we have

$$\begin{aligned} \text{pd}(\text{coker}\ \varphi_1)_Q &< k \\ &\leq \text{depth}\ I \\ &\leq \text{depth}\ I_Q \end{aligned}$$

which is a contradiction.

Finally, we prove that $\text{rad}\ I(\varphi_d) \neq \text{rad}\ I(\varphi_{d+1})$. We have already shown that $\text{rad}\ I(\varphi_d) = \text{rad}\ I(\varphi_1)$, so

$$\text{depth}\ I(\varphi_d) = \text{depth rad}\ I(\varphi_d) = \text{depth rad}\ I(\varphi_1) = d.$$

On the other hand, by Theorem 20.9, $I(\varphi_{d+1})$ must have depth $\geq d+1$ and we are done.

Corollary 20.12, in conjunction with Theorem 20.9, contains a well-known result of Auslander and Buchsbaum.

Corollary 20.13 (Auslander-Buchsbaum). *Let M be a module with a finite free resolution over a ring R. If the annihilator of M is nonzero, then it contains a nonzerodivisor, and the alternating sum of the ranks of the free modules in any finite free resolution of M (the "Euler characteristic of M") is zero.*

Proof. With notation as in Corollary 20.12, suppose $\mathcal{F}$ is a finite free resolution of M. By Theorem 20.9, $I(\varphi_1)$ contains a nonzerodivisor. By Proposition 20.7, the annihilator of M contains $\text{Fitt}_0\, M$. By Corollary 20.12,

$\operatorname{rank}\varphi_1 = \operatorname{rank} F_0$, so $\operatorname{Fitt}_0 M = I(\varphi_1)$ proving the first statement. The statement about the Euler characteristic follows from condition 1 of Theorem 20.9.

A beautiful application of Corollary 20.13, due to Vasconcelos, is given in Exercise 20.23.

The $I(\varphi_d)$ in a finite free resolution of a module M can be used to give an interesting description of the associated primes of M.

Corollary 20.14. *Let M be an R-module with a finite free resolution*

$$0 \to F_n \xrightarrow{\varphi_n} \cdots \to F_1 \xrightarrow{\varphi_1} F_0 \to M \to 0.$$

a. *If P is a prime of R and $d = \operatorname{depth} P_P$, then $P \in \operatorname{Ass} M$ iff $P \supset I(\varphi_d)$.*

b. *Set $\operatorname{depth} \operatorname{ann} M = d$. We have $\operatorname{depth} P_P = d$ for all $P \in \operatorname{Ass} M$ iff $\operatorname{depth} I(\varphi_k) > k$ for all $k > d$.*

Proof.

a. Corollary 19.10 of the Auslander-Buchsbaum formula, applied over the local ring R_P, shows that P is an associated prime of M iff the projective dimension of M_P over R_P is depth $P_P = d$. This happens iff $(\operatorname{coker} \varphi_d)_P$ is not free. By Proposition 20.8, $(\operatorname{coker} \varphi_d)_P$ is not free over R_P iff P contains $I((\varphi_d)_P)$. By Theorem 20.9, $I(\varphi_d)$ contains a nonzerodivisor, so $I((\varphi_d)_P) = I(\varphi_d)_P$, and we see that P is an associated prime of M iff P contains $I(\varphi_d)$, as required.

b. By Lemma 18.1, depth $I(\varphi_k) > k$ iff no associated prime P with depth $P_P = k$ contains $I(\varphi_k)$. Thus by part a, depth $I(\varphi_k) > k$ iff no such prime is associated to M. Since this is true for each $k > d$, the result follows. □

20.4 The Hilbert-Burch Theorem

In a sense Theorem 20.9 shows that the condition for a complex to be exact is not so complicated; but being a complex is a very subtle property! Thus, from Theorem 20.9, we are not really in a position to construct free resolutions, and in general it is very difficult to say what they look like: For example, it is an open conjecture of Buchsbaum and Eisenbud [1977], and Horrocks (reported in Hartshorne [1979]) that over a regular local ring, the Koszul complex has the smallest rank free modules of any resolution of a module of finite length.

To understand free resolutions further, one must look at some simple special cases. Let R be a local ring. In general, the most interesting modules to resolve are factor rings R/I, since in the geometric case these correspond

to subvarieties. If R/I has projective dimension 1, then I must be free and is thus generated by a nonzerodivisor.

The next case, where pd $R/I = 2$, is more interesting but can still be analyzed completely. This is the content of the **Hilbert-Burch theorem.**

Theorem 20.15.

a. *If a complex*
$$\mathcal{F} : 0 \to F_2 \xrightarrow{\varphi_2} F_1 \xrightarrow{\varphi_1} R \to R/I \to 0$$
is exact, and $F_1 \cong R^n$, *then* $F_2 \cong R^{n-1}$ *and there exists a nonzerodivisor* a *such that* $I = aI_{n-1}(\varphi_2)$. *In fact the* i*th entry of the matrix for* φ_1 *is* $(-1)^i a$ *times the minor obtained from* φ_2 *by leaving out the* i*th row. The ideal* $I_{n-1}(\varphi_2)$ *has depth exactly 2.*

b. *Conversely, given any* $(n-1)\times n$ *matrix* φ_2 *such that* depth $I(\varphi_2) \geq 2$, *and a nonzerodivisor* a, *the map* φ_1 *obtained as in part a makes* $\mathcal{F}$ *into a free resolution of* R/I, *with* $I = aI_{n-1}(\varphi_2)$.

Hilbert [1890] proved this result for graded ideals of codimension 2 in a polynomial ring, simply to give examples of free resolutions. Burch [1968] proved it in the generality here. Many other people have discovered it for themselves (and quite a few have also published it) in the intervening years. For a proof in a different style from that presented here see Kaplansky [1970, p. 148]. Our presentation reverses history, since the Hilbert-Burch theorem provided key evidence for the conjecture that became Theorem 20.9.

Proof.

a. To prove that φ_1 has the desired form, we choose bases for F_2 and F_1 and define $\varphi_1' : F_1 \to R$ to be the map with matrix whose ith entry is the minor obtained from φ_2 by leaving out the ith row. The composition $\varphi_1'\varphi_2$ is 0 because each of the entries of $\varphi_1'\varphi_2$ can be interpreted as the determinants of a matrix obtained from φ_2 by repeating a column. By the characterization of resolutions, $I_{n-1}(\varphi_2)$ has depth ≥ 2, so using the characterization again, we see that the second row in the following diagram is exact:

$$\begin{array}{lccccccc}
\mathcal{F}^* : & 0 & \to & R & \xrightarrow{\varphi_1^*} & F_1^* & \xrightarrow{\varphi_2^*} & F_2^* \\
 & & & {\scriptstyle a}\downarrow & & \downarrow & & \downarrow \\
\mathcal{F}'^* : & 0 & \to & R & \xrightarrow{\varphi_1'^*} & F_1^* & \xrightarrow{\varphi_2^*} & F_2^*
\end{array}$$

Here the two vertical maps on the right are the identity, and the map labelled a is a comparison map that exists because the first row is a free complex and the second row is exact. Since it is a map from R to R, it is multiplication by an element, which we shall also call a, of R.

It is evident from the commutativity of the diagram that $a\varphi_1' = \varphi_1$. It follows that a is a nonzerodivisor because $I(\varphi_1)$ has depth ≥ 1 by the characterization of resolutions.

To show that depth $I_{n-1}(\varphi_2) \leq 2$, we first apply Theorem 20.9 to $\mathcal{F}'$. The hypotheses are obviously fulfilled, and we see that $\mathcal{F}'$ is a resolution of $R/I_{n-1}(\varphi_2)$. Since $\mathcal{F}'$ has length 2, Corollary 18.5 shows that depth $I_{n-1}(\varphi_2) \leq \operatorname{pd}_R R/I_{n-1}(\varphi_2) = 2$. Alternatively, we could quote Exercise 10.9 to see that $\operatorname{codim} I_{n-1}(\varphi_2) \leq 2$ and use the fact that the depth is bounded above by the codimension (Proposition 18.2).

b. Part b follows immediately from Theorem 20.9. □

The Hilbert-Burch theorem is most often used in the study of Cohen-Macaulay rings of codimension 2—that is, Cohen-Macaulay factor rings $R = S/I$ of a regular ring S (usually taken to be local or graded) for which $\operatorname{codim} I = 2$. If R is such a factor ring, then since $\dim S = 2 + \dim R = 2 + \operatorname{depth} R$ in this case, we must have $\operatorname{pd}_S R = 2$ by the Auslander-Buchsbaum formula, and the Hilbert-Burch theorem applies. Conversely, if we are given an $n \times (n-1)$ matrix whose $(n-1) \times (n-1)$ minors generate an ideal I of codimension 2 in a regular ring S, then S/I will be Cohen-Macaulay of codimension 2. (This converse—but not the other implication—can be generalized to minors of other sizes and many types of related ideals, such as those defined by the Pfaffians of a skew-symmetric matrix.) An application of these ideas to factoriality of hypersurface rings is given in Exercise 20.17.

Perhaps the simplest setting where the Hilbert-Burch theorem is useful is in the study of ideals of sets of points in the projective plane. Indeed, any one-dimensional affine ring where every associated prime is minimal is Cohen-Macaulay, so that the ideal $I \subset S := k[x_0, x_1, x_2]$ of a set of points in $\mathbf{P}^2$—or indeed of any zero-dimensional scheme in $\mathbf{P}^2$—has $\operatorname{pd} S/I = 2$ by the Auslander-Buchsbaum formula. Thus the Hilbert-Burch theorem may be applied.

20.4.1 *Cubic Surfaces and Sextuples of Points in the Plane*

Here is an ancient example from projective geometry that still inspires research. See Gimigliano [1989] and Dolgachev and Kapranov [1993] for recent treatments and extensions. Let X be a set of six distinct points in $\mathbf{P}^2$, not contained in any conic. Since X does not lie on a conic, the ideal of X must be generated by forms of degree ≥ 3. The condition for a curve of given degree to pass through a given point is one linear condition on the coefficients of the equation of that curve; since the vector space of cubic equations has dimension 10, it follows that if J is the vector space of equations of cubics containing X, then

$$\dim J \geq 10 - 6 = 4.$$

Further, a minimal system of generators for the ideal I of X must contain a basis for J. Now by the Hilbert-Burch theorem, the minimal generators for I are the $(n-1)\times(n-1)$ minors of some $n\times(n-1)$ matrix with entries in the maximal homogeneous ideal (x_0, x_1, x_2) of S. In particular, if the ideal requires $\geq n$ generators, then the generators must each be of degree $\geq n-1$. In the present case, since we have cubic generators, I must be minimally generated by ≤ 4 elements, and we see that $\dim J = 4$ and any basis for J is a set of generators of I. Such a basis $f_1, \ldots, f_4$ is the set of 3×3 minors of a 3×4 matrix of forms of degrees ≥ 1, so in fact it must be the 3×3 minors of a 3×4 matrix M of linear forms. (More generally, the Hilbert-Burch theorem has the curious consequence that an ideal of a set of points with many generators must have all generators of rather high degree; this was the application that Burch originally had in mind.)

We can use the f_j to define a surface Y in $\mathbf{P}^3$, the image of $\mathbf{P}^2$ under the "rational map" f defined by sending a point $p\in\mathbf{P}^2$ to the point $(f_1(p), \ldots, f_4(p))\in\mathbf{P}^3$. The fiber of f over a general point is finite since, regarding the f_j as defining a map from $\mathbf{A}^3$ to $\mathbf{A}^4$, the preimage of the point 0 consists of the cone over X and is thus one-dimensional. By Corollary 14.5, the image of f is a surface in $\mathbf{P}^3$. Algebraically, Y may be defined as the surface whose equation is the generator of the kernel of the map of rings

$$k[y_0, \ldots, y_3] \to k[x_0, x_1, x_2] \text{ sending } y_j \text{ to } f_j(x).$$

The kernel is principal since it is a prime ideal of codimension 1, the ring generated by the f_j being a domain of dimension 3. In fact, the image is the plane, blown up at the "base locus" X of the four cubics.

We may compute the equation of this surface conveniently from the matrix M as follows. Restricting M to the degree $=0$ part of S^3, we get a map of vector spaces

$$k^3 = (S^3)_0 \to (S^4)_1 = \langle x_0, x_1, x_2\rangle \otimes_k k^4 = k^3 \otimes_k k^4.$$

In general, giving a $p\times q$ matrix of linear forms in n variables is the same as giving a map $k^q \to k^p \otimes k^n$. Thus the map $k^3 \to k^3\otimes k^4$ corresponding to M above also corresponds to a 3×3 matrix N of linear forms in four variables y_i. In fact, the determinant of this matrix, which is a cubic form in the y_i, is the equation of Y in $\mathbf{P}^3$. We shall not prove this here. However, it is easy to see at least that the determinant vanishes on Y; that is, substituting the f_i for the y_i in N gives a matrix $N(f)$ whose determinant is identically 0. Indeed the fact that $(f_1, \ldots, f_4)\circ M = 0$ is, as one easily sees, equivalent to the fact that $((x_0, x_1, x_2)\circ N)(f) = 0$.

20.5 Castelnuovo-Mumford Regularity

Throughout this section $S = k[x_1, \ldots, x_r]$ is a polynomial ring in r variables over a field k, and M denotes a finitely generated graded S-module. To simplify notation, we shall write Ext for Ext_S.

There is a notion of regularity for sheaves on projective spaces due to David Mumford [1966] that generalizes the idea of Castelnuovo's base-point free pencil trick given in Exercise 17.18; it is useful in controlling the vanishing of cohomology of a sheaf. A closely related notion for graded modules arises naturally in the study of finite free resolutions, and we present it here.

If $I \subset S$ is a homogeneous ideal, minimally generated by forms $f_1, \ldots, f_n$, then the maximum of the degrees of the f_i is an invariant of I—in fact it is even an invariant of the graded ring S/I, as the reader may easily check. Unfortunately, this invariant is difficult to handle. It is often easier to study the degrees of elements required to generate all the syzygies of I. To get a feeling for the right question, let M be any finitely generated graded S-module, and let

$$\cdots \to F_j \to \cdots \to F_0 \to M \to 0$$

be the minimal graded free resolution of M. Write b_j for the maximum of the degrees of the generators of F_j. In many cases of interest one has $b_j > b_{j-1} > \cdots > b_0$ (see Exercise 20.19), so it is natural to work with the sequence $b_j - j$. We shall say that M is ***m*-regular** for some integer m if $(b_j - j) \leq m$ for all j; and we define the **regularity** of M, written $\operatorname{reg} M$, to be the least integer m for which M is m-regular.

The definition is set up so that if M is m-regular, then in particular M is generated by elements of degrees $\leq m$. We may restate the definition by saying that the regularity of M is the smallest integer m such that for every j the jth syzygy of M is generated in degrees $\leq m + j$.

If M is m-regular, then F_j has no generators of degree $\geq m + j + 1$, so $F_j^* = \operatorname{Hom}_S(F_j, S)$ is zero in degree $\leq -m - j - 1$. Thus, in particular, $\operatorname{Ext}^j(M, S)$, which is a subquotient of F_j^*, must be 0 in degrees $\leq -m-j-1$. Not so obviously, the converse is true as well:

Proposition 20.16. *With notation as above, M is m-regular iff*

$$\operatorname{Ext}^j(M, S)_n = 0 \quad \textit{for all } j \textit{ and all } n \leq -m - j - 1.$$

Proof. We have already seen that if M is m-regular then the condition on Ext is satisfied. Suppose conversely that m is an integer for which the condition on Ext is satisfied. Let m' be the regularity of M. It suffices to show that $m' \leq m$.

Let j be the largest integer such that $b_j - j = m'$. The module $\operatorname{Ext}^j(M, S)$ is the homology of the dual of the resolution of M at F_j^*. From the hypothesis on j we see that F_j^* has $S(m' + j)$ as a summand, whereas F_{j+1}^* has no summand of the form $S(n)$ with $n \geq m' + j + 1$. Since the free resolution we have taken is minimal, the summand $S(m' + j)$ in F_j^* must map to 0 in F_{j+1}^*. Again by minimality (or by inspecting degrees) nothing in F_{j-1}^* can map onto the generator of $S(m' + j)$ in F_j^*, so it gives a nonzero class in

$\mathrm{Ext}^j(M,S)$, of degree $-m'-j$. Thus $-m'-j \geq -m-j$, so $m' \leq m$ as required.

Proposition 20.16 is hard to apply because in principle infinitely many conditions must be checked. Actually, it suffices to check just a few! To state the result cleanly we make another definition.

Definition. *We say that M is **weakly m-regular** if*

$$\mathrm{Ext}^j(M,R)_n = 0 \quad \textit{for all } j \textit{ and } n = -m-j-1.$$

In terms of this definition, Proposition 20.16 says that M is m-regular iff M is weakly m'-regular for all $m' \geq m$.

For most purposes this is the same as the definition for sheaves on projective space given by Mumford [1966]. (See the historical comments at the end of this section and also Exercise 20.20 for the translation.) As Mumford says, "This apparently silly definition reveals itself as follows:"

Theorem 20.17 (Mumford). *With notation as above, let N be the maximal submodule of M having finite length. If M is weakly m-regular, then M/N is m-regular.*

The context in which Mumford actually worked is that of sheaves over projective space. Every sheaf corresponds to a module (its module of twisted sections) having no submodule of finite length. In this context, Theorem 20.17 says simply that regularity coincides with what we have called weak regularity.

Proof. We do induction on the Krull dimension of the module M. Set $M' = M/N$. If $\dim M = 0$ then $M' = 0$, so the result is obvious.

From the exact sequence $0 \to N \to M \to M' \to 0$, we get a long exact sequence

$$\cdots \to \mathrm{Ext}^{j-1}(N,S) \to \mathrm{Ext}^j(M',S) \to \mathrm{Ext}^j(M,S) \to \mathrm{Ext}^j(N,S) \to \cdots$$

By Proposition 18.4 we have $\mathrm{Ext}^j(N,S) = 0$ for all $j < r$. Thus $\mathrm{Ext}^j(M',S) \cong \mathrm{Ext}^j(M,S)$ for $j < r$, so M' is also weakly m-regular.

The maximal homogeneous ideal is not associated to M'. We would like to choose a homogeneous element $x \in S$ of degree 1 such that x is a nonzerodivisor on M'. If k is infinite, this is no problem: Since the vector space of forms of degree 1 in S is not contained in any one of the associated primes of M', it is not contained in their union. If k is finite, there may be no such element x, but we can get around the problem as follows: Let K be an infinite extension field of k. Since K is flat over k, the minimal free resolution of $M \otimes_k K$ as a module over $S \otimes_k K$ is obtained by tensoring the resolution of M' as an S-module with K, and we see that $\mathrm{Ext}^j_{S\otimes_k K}(M' \otimes_k K, S \otimes_k K) = \mathrm{Ext}^j_S(M',S) \otimes_k K$. Thus neither the hypothesis nor the conclusion of Theorem 20.17 is affected by tensoring with K. It follows

that we may assume that k is infinite from the outset, and thus that a degree 1 nonzerodivisor x exists.

Let $\bar{M} = M'/xM'$. From the exact sequence $0 \to M'(-1) \xrightarrow{x} M' \to \bar{M} \to 0$, we get an exact sequence

$$(*)\quad \mathrm{Ext}^{j-1}(M'(-1), S) \to \mathrm{Ext}^j(\bar{M}, S) \to \mathrm{Ext}^j(M', S) \to \mathrm{Ext}^j(M'(-1), S).$$

It follows that $\bar{M}$ is weakly m-regular. If $\bar{N}$ is the largest submodule of finite length in $\bar{M}$, then $\bar{M}/\bar{N}$ is m-regular by induction on the dimension.

We can use the exact sequence $(*)$ to show that M' is weakly $(m+1)$-regular as follows: Since $\mathrm{Ext}^j(\bar{M}, S) = \mathrm{Ext}^j(\bar{M}/\bar{N}, S)$ for $j < r$, we see using Proposition 20.16 that $\mathrm{Ext}^j(\bar{M}, S)_n = 0$ for all $j < r$ and $n \le -m - j - 1$. For $n = -m - j - 2 = -(m+1) - j - 1$, we have

$$\begin{aligned}\mathrm{Ext}^j(M'(-1), S)_n &= \mathrm{Ext}^j(M', S)(1)_n\\ &= \mathrm{Ext}^j(M', S)_{n+1}\\ &= 0,\end{aligned}$$

since M' is weakly m-regular.

Thus by taking the degree $n = -m - j - 2$ part of the preceding exact sequence, we get $\mathrm{Ext}^j(M', S)_n = 0$ for $j < r$. Since $\mathrm{Ext}^j(M', S) = 0$ for $j = r$, this shows that M' is weakly $(m+1)$-regular. Applying this repeatedly, we see that M' is weakly m'-regular for all $m' \ge m$, so by Proposition 20.16 M' is m-regular. □

Putting these things together we get the following theorem.

Theorem 20.18. *Let $S = k[x_1, \ldots, x_r]$, with k a field, let M be a finitely generated graded S-module, and let N be the maximal submodule of M having finite length. The following conditions on an integer m are equivalent:*

a. M is m-regular.

b. M is weakly m-regular, and $N_t = 0$ for $t > m$.

Proof. Suppose that M is m-regular. By Proposition 20.16, M is weakly m-regular, so it suffices to prove the condition on N.

From the exact sequence $0 \to N \to M \to M/N \to 0$ we see that $\mathrm{Ext}^r(M, S) \cong \mathrm{Ext}^r(N, S)$. By Exercise 2.4 we have $\mathrm{Ext}^r_S(N, S(-r)) \cong \mathrm{Hom}_k(N, k)$. By Proposition 20.16, $\mathrm{Ext}^r_S(N, S)_n = 0$ for $n < -m - r$, so that $\mathrm{Ext}^r_S(N, S(-r))_n = 0$ for $n < -m$. Taking the dual into k, we see that $N_t = 0$ for $t > m$, as required.

For the converse, suppose M is weakly m-regular and $N_t = 0$ for $t > m$. By the previous duality argument it follows that $\mathrm{Ext}^r_S(M, S)_n = 0$ for $n < -m - r$. By Theorem 20.17 we know that M/N is n-regular for all $n \ge m$. Since $\mathrm{Ext}^j(M, S) \cong \mathrm{Ext}^j(M/N, S)$ for $j < r$, this shows that $\mathrm{Ext}^j(M, S)_n = 0$ for $n < -m - j$. By Proposition 20.16, M is m-regular.

As a first consequence we have:

Corollary 20.19. *If* $0 \to A \to B \to C \to 0$ *is a short exact sequence of graded finitely generated* S*-modules, then*

a. $\operatorname{reg} A \le \max(\operatorname{reg} B, \operatorname{reg} C + 1)$.

b. $\operatorname{reg} B \le \max(\operatorname{reg} A, \operatorname{reg} C)$.

c. $\operatorname{reg} C \le \max(\operatorname{reg} A - 1, \operatorname{reg} B)$.

d. If A *has finite length, then* $\operatorname{reg} B = \max(\operatorname{reg} A, \operatorname{reg} C)$.

Proof. We have a long exact sequence

$$\cdots \to \operatorname{Ext}^{j-1}(A,S) \to \\ \operatorname{Ext}^j(C,S) \to \operatorname{Ext}^j(B,S) \to \operatorname{Ext}^j(A,S) \to \operatorname{Ext}^{j+1}(C,S) \to \cdots,$$

from which conditions a–c follow at once, using Theorem 20.18.

If A has finite length, then $\operatorname{Ext}^j(A,S) = 0$ for all $j < r$, so the preceding long exact sequence degenerates into isomorphisms

$$\operatorname{Ext}^j(C,S) \cong \operatorname{Ext}^j(B,S) \quad \text{for } j < r,$$

and a short exact sequence

$$0 \to \operatorname{Ext}^r(C,S) \to \operatorname{Ext}^r(B,S) \to \operatorname{Ext}^r(A,S) \to 0.$$

Using the Theorem 20.18 characterization again, we see that both $\operatorname{reg} A$ and $\operatorname{reg} C \le \operatorname{reg} B$, and with part b we get the desired equality.

20.5.1 Regularity and Hyperplane Sections

One of the things that makes regularity a useful notion in the case of sheaves on projective space is that it is preserved when passing to a "general" hyperplane section. The algebraic expression of this is as follows.

Proposition 20.20. *If* M *is a finitely generated graded* S*-module and* x *is a linear form of* S *that is a nonzerodivisor on* M*, then* $\operatorname{reg} M = \operatorname{reg}(M/xM)$*. More generally, if* x *is a linear form whose annihilator* $(0 :_M x)$ *in* M *has finite length, then*

$$\operatorname{reg} M = \max(\operatorname{reg}(0 :_M x), \operatorname{reg} M/xM).$$

Proof. It follows directly from the definition that $\operatorname{reg} M(-1) = \operatorname{reg} M + 1$. If x is a nonzerodivisor, then $xM \cong M(-1)$, so $\operatorname{reg} xM = \operatorname{reg} M + 1$. Assuming only this equality, we shall show that $\operatorname{reg} M = \operatorname{reg} M/xM$.

Applying part a of Corollary 20.19 to the exact sequence

$$0 \to xM \to M \to M/xM \to 0,$$

we get $\operatorname{reg} M + 1 = \operatorname{reg} xM \leq \max(\operatorname{reg} M, \operatorname{reg} M/xM + 1)$, so $\operatorname{reg} M \leq \operatorname{reg} M/xM$. Applying part c of Corollary 20.19, we get $\operatorname{reg} M/xM \leq \max(\operatorname{reg} xM - 1, \operatorname{reg} M) = \operatorname{reg} M$, the opposite inequality.

In the general case, set $N = (0 :_M x)$ and suppose that N has finite length. The exact sequence $0 \to N(-1) \to M(-1) \to M \to M/xM \to 0$, where the middle map is multiplication by x, breaks into two short exact sequences

$$0 \to N(-1) \to M(-1) \to xM \to 0,$$
$$0 \to xM \to M \to M/xM \to 0.$$

Using part d of Corollary 20.19, we see from the first of these sequences that $\operatorname{reg} N \leq \operatorname{reg} M$ and $\operatorname{reg} xM \leq \operatorname{reg} M + 1$. From part c of Corollary 20.19 applied to the second sequence, we get $\operatorname{reg} M/xM \leq \max(\operatorname{reg} xM - 1, \operatorname{reg} M) = \operatorname{reg} M$, so $\operatorname{reg} M \geq \max(\operatorname{reg} N, \operatorname{reg} M/xM)$. It suffices to prove the opposite inequality.

Assume that $\operatorname{reg} N < \operatorname{reg} M$. Part d of Corollary 20.19 then shows that $\operatorname{reg} xM = \operatorname{reg} M(-1) = \operatorname{reg} M + 1$. By the argument in the first part of the proof, this implies $\operatorname{reg} M = \operatorname{reg} M/xM$, and we are done. □

20.5.2 Regularity of Generic Initial Ideals

Our final goal in this section is a result of Bayer and Stillman [1987a] connecting the regularity of an arbitrary ideal with the regularity of its generic initial ideal.

Corollary 20.21 (Bayer-Stillman). *Let $S = k[x_1, \ldots, x_r]$ be a polynomial ring over an infinite field k. Let F be a graded free S-module with basis and a reverse lexicographic monomial order. If $M \subset F$ is a graded submodule, then*

$$\operatorname{reg} F/M = \operatorname{reg} F/\operatorname{Gin}(M).$$

Proof. Rechoosing the variables $x_1, \ldots, x_r$ generically, we may assume in particular that $\operatorname{in}(M) = \operatorname{Gin}(M)$. Since the only associated prime of F/M containing x_r is the maximal ideal, we see that $(M : x_r)$ is of finite length. By Proposition 15.12b, we have $(\operatorname{in}(M) : x_r) = \operatorname{in}(M : x_r)$. Thus, $(\operatorname{in}(M) : x_r)/\operatorname{in}(M)$ and $(M : x_r)/M$ are nonzero in the same degrees, and it follows from Theorem 20.18 that they have the same regularity. By Noetherian induction we may assume that the result holds for $M + x_rF$, so we are done by Proposition 20.20.

20.5.3 Historical Notes on Regularity

The germ of the idea of regularity is present in work of Castelnuovo [1893] as (a special case of) the "base point free pencil trick" described in Exercise 17.18; see Exercise 20.21 for an explanation of the connection. This

was (according to Steve Kleiman) one of the things Oscar Zariski taught to all his students—including of course David Mumford. Mumford [1966] was the first to define the notion in essentially the generality given here. The approach we have followed is from Eisenbud and Goto [1984].

Mumford actually defines regularity for a coherent sheaf on projective space, using the vanishing of the cohomology of sheaves. His definition is equivalent to ours, applied to the module of twisted sections, under mild conditions; see Exercise 20.20. The idea was also exploited by Kleiman in his thesis [1965] and afterwards in some Exposés in Grothendieck's volume SGA 6 (Berthelot [1971]). In both of these references, the notion of regularity is used in the construction of bounded families of ideals or sheaves with given Hilbert polynomial—a crucial point in the construction of the Hilbert or Picard scheme. In a related direction, Kleiman [1971] uses the idea to prove that if I is a prime ideal in $S = k[x_1, \ldots, x_r]$ (and a little more generally), then the coefficients of the Hilbert polynomial of S/I can be bounded in terms of the initial coefficient alone.

20.6 Exercises

Exercise 20.1 (Generalized local rings; see Goto and Watanabe [1978]): We define a **generalized local ring** to be a positively graded ring $R = R_0 \oplus R_1 \oplus \cdots$ with R_0 local and Noetherian, and R finitely generated as an algebra over R_0. Show that R has a unique maximal homogeneous ideal, that we shall simply call the **maximal ideal** of R. Show that every finitely generated graded R-module has a unique minimal graded free resolution, just as in Theorem 20.2. Most of the results about local rings in this book can be adapted to generalized local rings, and include the analogous results in the local and graded cases.

Exercise 20.2: Let R be any commutative ring. Prove that an ideal $I \subset R$ is free as an R-module iff I is a principal ideal generated by a nonzerodivisor.

Fitting Ideals and the Structure of Modules

Exercise 20.3:*

a. (Fitting) Show that if R is a principal ideal domain, then any finitely generated R-module is determined up to isomorphism by its Fitting invariants. This fails for more complicated rings. For example, show that

b. If $R = k[x, y, z]$, with k a field, then the matrices

$$\begin{pmatrix} x & y \\ 0 & z \end{pmatrix} \text{ and } \begin{pmatrix} x & 0 \\ y & z \end{pmatrix}$$

present nonisomorphic modules with the same Fitting invariants.

c. Show that if we replace z by y in the matrices of part b, we get presentations of isomorphic modules; but if we replace z by y^2 we get presentations of nonisormorphic modules with the same Fitting invariants over $k[x, y]$.

Exercise 20.4: Extend Lemma 20.10 by showing that a complex

$$F \xrightarrow{\varphi} G \xrightarrow{\psi} H$$

of free R-modules with $I(\varphi) = R$ is exact iff $\operatorname{rank}\varphi + \operatorname{rank}\psi = \operatorname{rank} G$ and $I(\psi)$ contains a nonzerodivisor.

Exercise 20.5:* Show that the case of Corollary–Definition 20.4 in which F and F' are not supposed finitely generated (or free), and R is not Noetherian, follows from the case where R is Noetherian and F, F' are finitely generated and free.

Exercise 20.6 (Fitting invariant and annihilator for a square matrix): Suppose that M is presented by a square matrix φ whose determinant is a nonzerodivisor. Let ψ be the matrix of cofactors of φ. Using the expressions $\det(\varphi) \cdot I = \psi\varphi = \varphi\psi$, show that

$$\operatorname{ann} M = (\det(\varphi) : \operatorname{Fitt}_1 M) = (\operatorname{Fitt}_0 M : \operatorname{Fitt}_1 M).$$

If M is presented by a $p \times n$ matrix, with $p > n$ and depth $\operatorname{ann} M = p - n + 1$ (the largest possible value), then it can be shown that

$$\operatorname{ann} M = \operatorname{Fitt}_0 M.$$

For these and for formulas relating the other Fitting ideals to the annihilators of exterior powers of M, see Buchsbaum and Eisenbud [1977].

Exercise 20.7:* Compute the ideals $I(\varphi_k)$ for the Koszul complex of any sequence of elements, one of which is a nonzerodivisor.

Use the result and Theorem 20.9 to give another proof that $\operatorname{depth}((x_1, \ldots, x_n), R/J)$ can be determined if one knows which $H^i(K(x_1, \ldots, x_n) \otimes R/J)$ vanish.

Exercise 20.8: Let R be a ring and let M be an R-module. Suppose that $\varphi : F \to G$ is a map of free modules. Show that $\varphi \otimes M$ is a monomorphism iff $I_r\varphi$ contains a nonzerodivisor on M, where $r = \operatorname{rank} F$. (This can be extended to a version of the characterization of free resolutions that tells when $\mathcal{F} \otimes M$ is exact; see Buchsbaum and Eisenbud [1973].)

Exercise 20.9 (A stronger annihilator lemma, part a): Proposition 20.7a can be strengthened as follows: Let R be a ring and let M be a finitely generated R-module. For each $j \geq 0$ let $A_j(M)$ be the annihilator of $\wedge^{j+1} M$. Thus $A_0(M) = \operatorname{ann} M$, and Proposition 20.7a says that

$\mathrm{Fitt}_0(M) \subset A_0(M)$. Strengthen this result by proving

(a') $\qquad\qquad \mathrm{Fitt}_j(M) \subset A_j(M) \quad \text{for all } j \geq 0$

as follows:

Suppose that $M = \operatorname{coker} \varphi : F \to G$ with F and G free and rank $G = n$. Let $\varphi_{i,j} : \wedge^i F \otimes \wedge^j G \to \wedge^{i+j} G$ be the map $f \otimes g \mapsto \wedge^i \varphi(f) \wedge g$ induced by φ. Set $A_{i,j} = \operatorname{ann} \operatorname{coker} \varphi_{i,j}$.

i. Show that $\mathrm{Fitt}_j(M) = A_{n-j,j}$ while $A_j(M) = A_{1,j}$ (for the latter, see Proposition A2.2d). Thus it suffices to show that for all $i + j \leq n$ we have

$(*)$ $$A_{i,j} \subset A_{i-1,j}.$$

ii. Now suppose that $g \in \wedge^{i+j-1} G$ and $r \in A_{i,j}$. To prove $(*)$ we must show that $rg \in \operatorname{im} \varphi_{i-1,j}$. By linearity, it suffices to prove this for all vectors g in some fixed basis for $\wedge^{i+j-1} G$, so we may suppose that $g = e_1 \wedge \cdots \wedge e_{i+j-1}$, where the e_i form a basis of G. Using this representation show that there is an element $g' \in \wedge^{i+j} G$ and a $\gamma \in G^*$ such that $\delta_\gamma(g') = g$, where δ_γ is the differential of the Koszul complex corresponding to $\gamma : G \to R$ as in Chapter 17. (In the appendix on multilinear algebra, this would be written simply as $\gamma(g') = g$.)

iii. By hypothesis there is an element $h = \sum f_i \otimes g_i \in \wedge^i F \otimes \wedge^j G$ such that $rg' = \sum (\wedge^i \varphi(f_i)) \wedge g_i$. Use the fact that δ_γ acts as a derivation (Proposition A2.8) to show that

$$rg = \delta_\gamma(rg') = \sum \delta_\gamma(\wedge^i \varphi(f_i)) \wedge g_i \pm \sum (\wedge^i \varphi(f_i)) \wedge \delta_\gamma g_i \in \wedge^{i+j-1} G.$$

Show from the definitions that $\delta_\gamma(\wedge^i \varphi(f_i)) = (\wedge^i \varphi(\delta_{\varphi^*(\gamma)} f_i))$ (this says that $\wedge \varphi$ is a map of complexes). The left-hand term is in the image of $\varphi_{i-1,j}$, and the right-hand term is in the image of $\varphi_{i,j-1}$.

iv. Complete the proof by showing that $\operatorname{im}(\varphi_{i,j-1}) \subset \operatorname{im}(\varphi_{i-1,j})$. (Note that the image of $\varphi_{i,j}$ is generated by all wedge products of i elements of $\varphi(F)$ and j arbitrary elements of G.)

Exercise 20.10 (A stronger annihilator lemma, part b): With notation as in Exercise 20.9, Proposition 20.7b says that for every $j > 0$ we have $A_0(M) \mathrm{Fitt}_j(M) \subset \mathrm{Fitt}_{j-1}(M)$. Prove a stronger result,

(b') $\quad$ For every $j > i \geq 0$ we have $A_i(M) \mathrm{Fitt}_j(M) \subset \mathrm{Fitt}_{j-1}(M)$,

by checking the following steps:

i. The image of $\varphi_{1,i-1} : F \otimes \wedge^{i-1} G \to \wedge^i G$ contains $A_i(M) \wedge^i G$. (See Proposition A2.2d.)

ii. Therefore the image of $\varphi_{1,j-1} : F \otimes \wedge^{j-1} G \to \wedge^j G$ contains $A_i(M) \wedge^j G$.

iii. The image of $\varphi_{1,i} : \wedge^{n-j} F \otimes \wedge^j G \to \wedge^n G$ is $\operatorname{Fitt}_j(M) \wedge^n G$.

iv. Therefore the image of $\wedge^{n-j} F \otimes F \otimes \wedge^{j-1} G \to \wedge^n G$ under the map $f_1 \otimes f_2 \otimes g \mapsto \wedge^{n-j}\varphi(f_1) \wedge \varphi(f_2) \wedge g$ has image containing $A_i(M) \operatorname{Fitt}_j(M) \wedge^n G$.

v. Complete the proof by observing that the image of the map in part iv is the same as the image of $\varphi_{n-j+1,j-1}$, which is $\operatorname{Fitt}_{j-1}(M)$.

Deduce from this exercise and the last that if M can be generated by n elements, then $A_j(M)^{n-j+1} \subset \operatorname{Fitt}_j(M)$, so that Radical $(\operatorname{Fitt}_j(M)) =$ Radical $(A_j(M))$. (There is an easy localization argument that proves this last equality; can the reader find it?) These and similar results are contained in Buchsbaum and Eisenbud [1977]. (Note that the Fitting ideals are indexed differently there than in this book, and the inequality in Theorem 1.2.2 in that paper points in the wrong direction.)

Projectives of Constant Rank

The following problems explain what it means for a projective module to have constant rank.

Exercise 20.11:* If M is a finitely generated R-module, show that the set

$$\{P \text{ a prime of } R | M_P \text{ can be generated by } \leq k \text{ elements over } R_P\}$$

is an open set of $\operatorname{Spec} R$.

Exercise 20.12:* Prove that every finitely generated projective R-module M has constant rank iff R has no nontrivial idempotents.

Exercise 20.13 (Fiberwise characterization of projectives):* Suppose that R is a reduced ring, and that M is a finitely generated R-module.

a. Show that M is projective iff the number

$$\mu_P(M) := \dim_{R_P/PR_P}(M_P/PM_P)$$

is a locally constant function of the primes P in R. Show that M is projective of constant rank iff this function is constant.

b. The hypothesis that R is reduced is necessary. Show that either of the statements in part a is a characterization—perhaps a silly one—of reduced rings.

Exercise 20.14 (Flatness and Hilbert functions): As in Exercise 6.11, let $R = R_0 \oplus R_1 \oplus \cdots$ be a graded ring. Suppose that R_0 is a Noetherian ring and that R is finitely generated as an R_0-algebra by elements of degree

1. Let M be a finitely generated graded R-module, and let M_d be the degree-d part of M. Thinking of M as a family of graded modules over the "base" R_0, we ask, for each prime ideal $P \subset R_0$, about the Hilbert function $H_{\kappa(P)\otimes M}(d) = \dim_{\kappa(P)} \kappa(P) \otimes M_d$ and about the associated Hilbert polynomial. Exercise 6.11d shows that if M is flat and R_0 is local, then $H_{\kappa(P)\otimes M}(d)$ is constant as a function of P. The same goes for the Hilbert polynomial $P_{\kappa(P)\otimes M}(d)$ under the weaker assumption that $M[f^{-1}]_0$ is flat over R_0 for all $f \in R_1$.

   a. Deduce from these local statements that for an arbitrary Noetherian R_0, if M as above is flat over R_0, then the function $H_{\kappa(P)\otimes M}(d)$ is locally constant as a function of P, and similarly for $P_{\kappa(P)\otimes M}(d)$. In particular, the dimension of the fiber $\kappa(P) \otimes_R M$ is locally constant as a function of P (in the Zariski topology on the set of primes of R).

   b. Using Exercises 20.13 and 6.11c, show that the converse is true when R_0 is reduced: That is, if $H_{\kappa(P)\otimes M}(d)$ is locally constant as a function of P, then M is flat over R_0, while if $P_{\kappa(P)\otimes M}(d)$ is locally constant as a function of P, then $M[f^{-1}]_0$ is flat over R_0 for all $f \in R_1$.

Exercise 20.15: Prove that a ring R is a factorial domain iff for every x, y in R, one of the two following conditions holds:

i. (x, y) is generated by a nonzerodivisor (possibly a unit!).

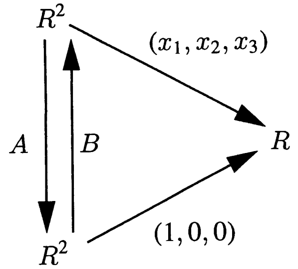
ii. There is a regular sequence x', y' in R and a nonzerodivisor a in R such that $x = ax', y = ay'$.

This can be used to show that a local ring is factorial iff every two-generator ideal has finite projective dimension (MacCrae's [1965] generalization of the Auslander-Buchsbaum theorem; see Buchsbaum and Eisenbud [1974] for a generalization replacing (x, y) by any matrix).

Exercise 20.16 (A nonunique resolution): The question of the uniqueness of free resolutions in the nonlocal case is quite subtle. Here is an example based on the example of the "tangent bundle of the 2-sphere," discussed in Chapter 19. Let

$$R = k[x_1, x_2, x_3]/(x_1^2 + x_2^2 + x_3^2 - 1)$$

where k is a field. Show that there are matrices A and B making the following diagram commute:
If $k \subset \mathbf{R}$, the field of real numbers, show that neither A nor B can be chosen to be an isomorphism. Thus resolutions of the ideal $(1) = (x, y, z) \subset R$ that begin with the upper and lower diagonal maps are not isomorphic. If $k = \mathbf{C}$, on the other hand, show that A and B can be taken to be isomorphisms. (There are similar examples, using a few more variables, where the resolutions remain non-isomorphic over the complex numbers.)



This example can be used to construct others where the ideals resolved are minimally generated by the given generators. Here is the "generic" example: Let T be a polynomial ring S in 24 variables $\{a_{ij}, b_{ij}, x_i, y_i\}_{1\leq i,j\leq 3}$ over k modulo to the 6 quadratic relations that give the matrix equalities

$$(x_1, x_2, x_3)A = (y_1, y_2, y_3) \quad \text{and} \quad (y_1, y_2, y_3)B = (x_1, x_2, x_3),$$

where A and B are the matrices with entries $a_{i,j}$ and b_{ij} respectively. Show that $(x_1, x_2, x_3)T = (y_1, y_2, y_3)T$. Construct the first terms of two T-free resolutions of this ideal, and show they are non-isomorphic in the sense above if $k = \mathbf{R}$.

Exercise 20.17 (Factoriality of Hypersurface rings): There are a number of beautiful results on the question of when a ring of the form $R/(f)$, where R is regular local (or a polynomial ring), is factorial. Perhaps the most famous is the Noether-Lefschetz theorem: If $R = \mathbf{C}[x_1, \ldots, x_4]$ is the polynomial ring in four variables over $\mathbf{C}$, then for almost every homogeneous form f of degree $\geq 4, R/(f)$ is factorial. Here the "almost every" *must* be taken in an analytic sense: The set of "bad" f is a countable union of hypersurfaces (and is thus not an algebraic set). For a modern treatment see Ciliberto, Harris, and Miranda [1988].

In a different direction, a result due to Brieskorn in characteristic 0 and to Lipman in greater generality says, in the simplest case, that if k is an algebraically closed field of characteristic 0 and $A := k[[x, y, z]]/(f)$ is factorial, then $A \cong k[[x, y, z]]/(x^2 + y^3 + z^5)$, the so-called E8 singularity. See Lipman [1975].

Here, as an application of the Hilbert-Burch theorem, we give a rather general result due to Andreotti and Salmon [1957], from which Brieskorn's result may be deduced (Choi [1988]). Let (R, P) be a three-dimensional regular local ring, and let $0 \neq f \in P$ be any element. Show that the ring $R/(f)$ is factorial iff f cannot be written as the determinant of an $n \times n$ matrix with entries in P for $n > 1$, as follows.

If $f = \det(a_{ij})$, let M be an $(n-1) \times n$ submatrix of (a_{ij}). Show that the ideal of minors of M is an unmixed ideal of codimension 2 in R, and therefore codimension 1 in $R/(f)$, that is not principal, so $R/(f)$ is not factorial.

If $R/(f)$ is not factorial, let Q be a prime of R of codimension 2 such that $f \in Q$ but $Q/(f)$ is not principal. Show that R/Q is Cohen-Macaulay, so that by Hilbert-Burch Q is minimally generated by the $(n-1)\times(n-1)$ minors of some $(n-1)\times n$ matrix M with entries in the maximal ideal. Show that f is the determinant of a matrix made by augmenting the matrix M with one further row to get an $n \times n$ matrix M'. If $f \in PQ$, the new matrix can be chosen with entries in P, and we are done. If $f \notin PQ$, show that $n \geq 3$, and that after row and column operations M' can be written as the direct sum of a 1×1 identity matrix and an $(n-1)\times(n-1)$ matrix with entries in P.

Castelnuovo-Mumford Regularity

Exercise 20.18: Show that if M has finite length then $\operatorname{reg} M = \max\{n | M_n \neq 0\}$.

Exercise 20.19: Let $F : \cdots \to F_j \to \cdots \to F_0 \to M \to 0$ be a minimal free resolution of a finitely generated graded module over the graded polynomial ring $S = k[x_1, \ldots, x_r]$. Write $F_j = \oplus_m S(-a_{j,m})$. Set $a_j = \min_m(a_{j,m})$, and $b_j = \max_m(a_{j,m})$.

Show that $a_j > a_{j-1} > \cdots > a_0$. If the dual $F^* : F_0^* \to F_1^* \to \cdots \to F_j^* \to \cdots$ is also exact, show that $b_j > b_{j-1} > \cdots > b_0$. Show that this condition is satisfied, in particular, when M is a module of finite length, or more generally when depth $M = \dim M$.

Exercise 20.20 (Regularity of sheaves on $\mathbf{P}^n$):* (For those who know something about cohomology of sheaves, and local duality.) Suppose that $\mathcal{F}$ is a sheaf on $\mathbf{P}^n$, and let $M := \oplus_d H^0(\mathbf{P}^n; \mathcal{F}(d))$ be the corresponding graded module of twisted sections. Mumford defines $\mathcal{F}$ to be m-regular if $H^j(\mathbf{P}^n; \mathcal{F}(m-j)) = 0$ for all $j > 0$.

a. Show that Mumford's definition is the same as saying that

$$\operatorname{Ext}^j(M, S)_n = 0 \quad \text{for all } j \leq r-2 \text{ and } n = -m-j-1.$$

b. Suppose that $\operatorname{Ass}(\mathcal{F})$ contains no closed points, so that M is finitely generated. Show that M is m-regular in the sense of this chapter iff $\mathcal{F}$ is m-regular in Mumford's sense.

Exercise 20.21 (Castelnuovo's base point free pencil trick revisited): Suppose that $\mathcal{L}$ is a line bundle on a curve X in $\mathbf{P}^n$, and that $\mathcal{L}$ is m-regular in the sense that $H^1\mathcal{L}(m-1) = 0$. Let $R = \oplus H^0(\mathcal{O}_X(d))$ be the homogeneous coordinate ring of X. Show using the previous exercise that the R-module $\oplus_n H^0\mathcal{L}(d)$ is generated in degree $\leq m$. (You may also deduce this from Exercise 17.18.)

Exercise 20.22 (Linear free resolutions): We say that a module M, generated in some degree n, has a **linear free resolution** if it has a free

resolution of the form

$$\cdots \to \oplus S(-n-j)^{\beta_j} \to \cdots \to \oplus S(-n-1)^{\beta_1} \to \oplus S(-n)^{\beta_0} \to M \to 0.$$

a. Suppose M is a module of finite length. Show that M has a linear free resolution iff $M \cong k(n)^\beta$ for some n and β—that is, iff M is a vector space concentrated in a single degree.

b. Suppose M is an ideal of S, and that S/M has finite length. Show that M has a linear free resolution iff M is a power of the maximal ideal.

c. Despite these cases, there are a lot of modules with linear free resolutions: Show that reg $M = m$ iff the "truncation" $T_m(M) := \oplus_{n \geq m} M_n$ has a linear free resolution.

See Eisenbud and Goto [1984] and Eisenbud and Koh [1991] for some further information and conjectures on this topic.

Exercise 20.23 (Vasconcelos' characterization of ideals generated by regular sequences): Suppose that (R, P) is a local ring and $I \subset R$ is an ideal. If I is generated by a regular sequence, then I/I^2 is a free module over R/I by Exercise 17.16, and I has finite projective dimension by Corollary 19.3. Prove that the converse is also true, as follows:

a. Since I has finite projective dimension, so does R/I. Thus I is the annihilator of a module of finite projective dimension, so by Corollary 20.13, I contains a nonzerodivisor. Using this and prime avoidance, show that there is a nonzerodivisor $x \in I$ such that $x \notin PI$.

b. Note that x generates an R/I-free summand of I/I^2 so that $I/I^2 \cong ((x) + I^2)/I^2 \oplus J/I^2$ for some ideal $J \supset I^2$ and $((x) + I^2)/I^2 \cong R/I$. Show that $(x) \cap J = xI$.

c. Show that $I/xI \cong I/(x) \oplus (x)/xI$.

d. Show that I/xI is a module of finite projective dimension over $R/(x)$. Conclude that $I/(x) \subset R/(x)$ is an ideal of finite projective dimension.

e. Use induction on the rank of the free R/I-module I/I^2 to finish the proof.

Derive from this characterization another proof that if R, P is a local ring of finite global dimension, then P is generated by a regular sequence, so that R is regular. For a refinement and further applications, see the original paper of Vasconcelos [1967].

21

Duality, Canonical Modules, and Gorenstein Rings

Throughout this chapter, we shall assume that all the rings considered are Noetherian.

The most basic geometric objects associated with a smooth manifold or a smooth variety are its tangent and cotangent bundles, and various tensors derived from them. In algebraic geometry the most easily accessible, and the most important, is the **canonical bundle**, the highest exterior power of the cotangent bundle. It plays a central part in duality theory. If the variety is affine, then the sections of the canonical bundle form a module over the coordinate ring of the variety called the **canonical module**. Because it is the module of sections of a line bundle, it is locally free of rank 1.

If the variety has singularities, then there is still a canonical module, but it may not be locally free. It is an interesting object and plays a major role in the theory of Cohen-Macaulay rings. In this chapter we shall introduce the local theory of the canonical module over a Cohen-Macaulay ring, and in particular we shall study local Gorenstein rings—those local rings for which the canonical module is free. What we do here is only a beginning. For a more extensive treatment, see Bass [1963], Herzog and Kunz [1971], or Bruns and Herzog [1993].

The global study of the properties of the canonical module is quite subtle. Despite a monumental attempt by Grothendieck and Hartshorne (Hartshorne [1966b]) and much work by others (see, for example, Kunz and Waldi [1988] and Lipman [1984]), I suspect that the subject is probably still not in its final form. In any case, its study is best undertaken in

the general context of schemes. The most accessible introduction seems to me to be Altman and Kleiman [1970].

We begin this chapter with a leisurely introduction to the zero-dimensional case, where many of the features of the theory can be clearly seen without the homological technique necessary for the general case, which we treat in the rest of the chapter. The reader who has not encountered injective modules and resolutions before will need to master these notions and the notion of essential extension before reading the details of the proofs of the later sections of this chapter. We give the necessary results in Appendix 3.

Duality theory, even in geometry, is a notably technical and algebraic matter. Most of this chapter is resolutely algebraic, but at the end we have presented some of the elements of the theory of liaison (linkage) and (in the exercises) Cayley-Bacharach theory to show that this material has remarkable geometric applications.

21.1 Duality for Modules of Finite Length

We begin with a very simple case. Suppose that k is a field and that A is a local zero-dimensional ring that is a finite-dimensional k-algebra. (See Exercise 21.3 for the nonlocal case.) If we wish to imitate the usual duality theory for vector spaces, we might at first try to work with the functor

$$M \mapsto \mathrm{Hom}_A(M, A).$$

But this is often very badly behaved; for example, it does not usually preserve exact sequences, and if we do it twice we do not get the identity: That is, $\mathrm{Hom}_A(\mathrm{Hom}_A(M, A), A) \not\cong M$ in general.

A good duality may be defined in a different way: If M is a finitely generated A-module, we provisionally define the dual of M to be

$$D(M) = \mathrm{Hom}_k(M, k)$$

(we shall give a more intrinsic definition shortly). The vector space $D(M)$ is naturally an A-module by the action

$$(a \cdot \varphi)(m) = \varphi(am) \quad \text{for } \varphi \in D(M), a \in A, \text{ and } m \in M.$$

This makes D a contravariant functor from the category of finitely generated A-modules to itself. Since M is finite-dimensional over k, the natural map $M \to D(D(M))$ sending $m \in M$ to the functional $\varphi \mapsto \varphi(m)$, for $\varphi \in \mathrm{Hom}_k(M, k)$, is an isomorphism of vector spaces; in fact it is an isomorphism of A-modules, as the reader may immediately check. Since k is a field, D is **exact** in the sense that it takes exact sequences to exact

sequences (with the arrows reversed). Thus D is a **dualizing functor**[1] on the category of finitely generated A-modules. This dualizing functor is very familiar in the theory of representations of groups, where $D(M)$ is called the **contragredient** representation to M.

The functor D seems to depend on the field k, but we shall show that it does not. For a transparent special case, suppose that A is a field, finite over the field k. For any vector space V over $A, D(V) = \mathrm{Hom}_k(V,k)$ is again a vector space over A. To show that $D(V) \cong \mathrm{Hom}_A(V,A)$, it suffices to prove that $\dim_A V = \dim_A D(V)$. This is immediate because $\dim_k V = \dim_k D(V)$. We shall soon see that there is even a functorial isomorphism. However, there is no canonical isomorphism: If there were, there would have to be a canonical isomorphism $A \cong \mathrm{Hom}_k(A,k)$, or equivalently a canonical choice of generator of this one-dimensional A-vector space. If A is separable over k, one may take the trace map, but if A is inseparable there seems to be no canonical choice.

To get an idea of how D acts, note first that if P is a maximal ideal of A, then any dualizing functor D takes the simple module A/P to itself. Indeed, $D(A/P)$ must be simple, because else it would have a proper factor module M and $D(M)$ would be a proper submodule of A/P. As A is local, it has only one simple module, and thus $D(A/P) \cong A/P$. Since D takes exact sequences to exact sequences, reversing the arrows, D "turns composition series upside down" in the sense that if

$$0 \hookrightarrow M_1 \hookrightarrow \cdots \hookrightarrow M_n \hookrightarrow M$$

is a chain of modules with simple quotients $M_i/M_{i-1} \cong A/P$, then

$$D(M) \twoheadrightarrow D(M_n) \twoheadrightarrow \cdots \twoheadrightarrow D(M_1) \twoheadrightarrow D(0) = 0$$

is a chain of surjections whose kernels N_i are simple. In particular, for any module of finite length, the length of $D(M)$ equals the length of M. More generally, the lattice of submodules of $D(M)$ is obtained from the lattice of submodules of M by reversing inclusions.

It is easy to see that any dualizing functor preserves annihilators: If $a \in A$ annihilates M, then by the A-linearity of D, a annihilates $D(M)$, so $\mathrm{ann}(M) \subset \mathrm{ann}(D(M))$. Since $D(D(M)) = M$, the reverse inclusion holds as well.

Any dualizing functor preserves endomorphism rings; more generally, $\mathrm{Hom}_A(D(M), D(N)) \cong \mathrm{Hom}_A(N, M)$. In particular, $D(A)$ is a module with endomorphism ring A. To see this consider the mappings given by applying D:

[1]Technical definition: A dualizing functor is a contravariant A-linear functor D such that $D^2 = 1$ and D is **exact** in the sense that D takes exact sequences to exact sequences; actually one can omit the condition of exactness—see Exercise 21.2.

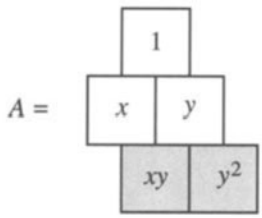
$$\begin{aligned}\operatorname{Hom}_A(M,M) &\to \operatorname{Hom}_A(D(M),D(M))\\ &\to \operatorname{Hom}_A(M,M) \to \operatorname{Hom}_A(D(M),D(M)).\end{aligned}$$

Since $D^2 \cong 1$, the composite of two successive maps in this sequence is an isomorphism, so each of the maps is an isomorphism too.

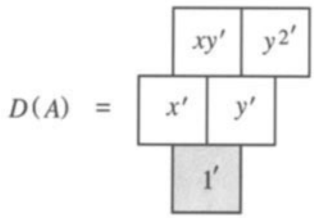
Some of the standard concepts of module theory have interesting duals. For example, the definition of an injective module may be obtained from the definition of a projective module by reversing the directions of the arrows; thus the dual of a projective module is an injective module, and similarly the dual of an injective module is projective.

A central role in the theory of modules over a local ring A, P is played by what might be thought of as the **top** of a module M, defined to be the quotient $M \twoheadrightarrow M/PM$; Nakayama's lemma shows that this quotient controls the generators of M. It could be defined categorically as the largest quotient of M that is a direct sum of simple modules. The dual notion is that of the **socle** of M: It is defined as the annihilator in M of the maximal ideal P, or equivalently, as the sum of all the simple submodules of M. (The word *socle* is an architectural term for the base of a column; a glance at the illustrations will explain its use in our context.) Note that since the top of A is A/P, a simple module, the socle of $D(A)$ must be a simple module as well. The socle of a graded module is naturally a graded vector space, but it need not consist of elements all of the same degree; see Exercise 21.1.

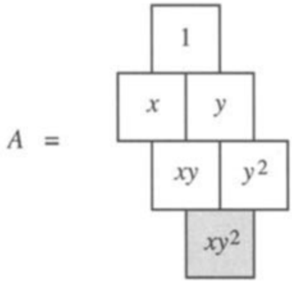
For example, if we picture the ring $A = k[x,y]/(x^2, xy^2, y^3)$ in terms of a basis of monomials as



then the dual module ω_A, which is injective as an A-module, might be pictured as follows:



Here for each monomial basis vector f, we have written f' for the dual basis vector to f. In each of these pictures we have shaded the boxes representing the socle. Note that A has a simple top, and $D(A)$ has a simple socle, as claimed. For the ring $A = k[x,y]/(x^2, y^3)$, the picture is symmetric:



We shall see that such symmetry holds for all complete intersections, and somewhat more generally.

The definition of D that we have given depends on the field k chosen, but in fact D does not!

Proposition–Definition 21.1. *Let (A,P) be a local zero-dimensional ring. If E is any dualizing functor from the category of finitely generated A-modules to itself, then there is an isomorphism of functors $E(-) \cong \mathrm{Hom}_A(-, E(A))$. Further, $E(A)$ is isomorphic to the injective hull of A/P. Thus there is up to isomorphism at most one dualizing functor.*

For the definitions of injective hull and essential submodule, used in the following proof, see Appendix 3.

Proof. Since $E^2 \cong 1$ as functors, the map $\mathrm{Hom}_A(M,N) \to \mathrm{Hom}_A(E(N), E(M))$ given by $\varphi \mapsto E(\varphi)$ is an isomorphism. Thus there is an isomorphism, functorial in M,

$$E(M) = \mathrm{Hom}_A(A, E(M)) \cong \mathrm{Hom}_A(E(E(M)), E(A)) \cong \mathrm{Hom}_A(M, E(A)).$$

This proves the first statement.

Since A is projective, $E(A)$ is injective. As we observed above, A has a simple top, so $E(A)$ has a simple socle. Because A is zero-dimensional, every module contains simple submodules. The socle of a module M contains all the simple submodules of M, and thus meets every submodule of M; that is, it is an essential submodule of M. Since A/P appears as an essential submodule of $E(A)$, we see that $E(A)$ is an injective hull of A/P.

With Proposition 21.1 for justification, we define the **canonical module** ω_A of a local zero-dimensional ring A (not necessarily containing a field) to be the injective hull of the residue class field of A. By Proposition 21.1 any **dualizing functor** D on the category of finitely generated A-modules must be $D(M) := \mathrm{Hom}_A(M, \omega_A)$, and in fact this functor is always dualizing.

Proposition 21.2. *If A is a zero-dimensional local ring, then the functor $M \mapsto D(M) := \operatorname{Hom}_A(M, \omega_A)$ is a dualizing functor on the category of finitely generated A-modules.*

Proof. The functor D is obviously A-linear, and it is exact because ω_A is injective. Thus it suffices to show that D^2 is isomorphic to the identity. Let $\alpha : 1 \to D^2$ be the natural transformation given by maps

$$\alpha_M : M \to \operatorname{Hom}_A(\operatorname{Hom}_A(M, \omega_A), \omega_A),$$

sending $m \in M$ to the homomorphism taking $\varphi \in \operatorname{Hom}_A(M, \omega_A)$ to $\varphi(m)$. We shall show that α is an isomorphism by showing that each α_M is an isomorphism.

We do induction on the length of M. First suppose that the length is 1, so that $M = A/P$, where P is the maximal ideal of A. Since ω_A is the injective hull of A/P, the socle of ω_A is A/P, and we have $\operatorname{Hom}_A(A/P, \omega_A) = A/P$, generated by any nonzero map $A/P \to \omega_A$. Thus $\operatorname{Hom}_A(\operatorname{Hom}_A(A/P, \omega_A), \omega_A) = A/P$, generated by any nonzero map. But if $1 \in A/P$ is the identity, then the map induced by 1 takes the inclusion $A/P \hookrightarrow \omega_A$ to the image of 1 under that inclusion, and is thus nonzero, so $\alpha_{A/P}$ is an isomorphism.

If the length of M is greater than 1, let M' be any proper submodule and let $M'' = M/M'$. By the naturality of α and the exactness of D^2 it follows that there is a commutative diagram with exact rows

$$\begin{array}{ccccccccc} 0 & \to & M' & \to & M & \to & M'' & \to & 0 \\ & & \downarrow{\scriptstyle \alpha_{M'}} & & \downarrow{\scriptstyle \alpha_M} & & \downarrow{\scriptstyle \alpha_{M''}} & & \\ 0 & \to & D^2M' & \to & D^2M & \to & D^2M'' & \to & 0. \end{array}$$

Both M' and M'' have lengths $<$ length M, so the left- and right-hand vertical maps are isomorphisms by induction. It follows by an easy diagram chase (see the "Five Lemma," Exercise A3.11) that the middle map α_M is an isomorphism too.

Corollary 21.3. *If A is a local Artinian ring, then the annihilator of ω_A is 0; the length of ω_A is the same as the length of A; and the endomorphism ring of ω_A is A.*

Proof. The dualizing functor preserves annihilators, lengths, and endomorphism rings, and takes A to ω_A.

If a zero-dimensional local ring A is a finite-dimensional vector space over a field k, then the canonical module may be expressed in terms of the field: A dualizing functor is $\operatorname{Hom}_k(-, k)$, so we have $\omega_A \cong \operatorname{Hom}_k(A, k)$. We next show that with a small modification we can replace the field by an arbitrary local ring. Later we shall give other formulas of this type.

Proposition 21.4. *Let A be a zero-dimensional local ring. Suppose that for some local ring B, A is a B-algebra that is finitely generated as a B-module and the maximal ideal of B maps into that of A. If E is the injective hull of the residue class field of B, then*

$$\omega_A = \mathrm{Hom}_B(A, E).$$

In particular, if B is also zero-dimensional, then $\omega_A = \mathrm{Hom}_B(A, \omega_B)$.

Proof. By Appendix A3.8, $\mathrm{Hom}_B(A, E)$ is an injective A-module. To show that it is the injective hull of the residue class field k of A, it suffices to show that it is an essential extension of the residue class field k of A. Let P be the maximal ideal of A, and let P_B be the maximal ideal of B. By Proposition 9.2, the preimage of P is P_B, so there is an induced homomorphism of the residue class field k_B of B to k. As k is a finite-dimensional vector space over k_B, we have $k = \omega_k \cong \mathrm{Hom}_{k_B}(k, k_B)$ as k-modules (this is in fact the special case of the proposition where A and B are fields).

Let $S \subset \mathrm{Hom}_B(A, E)$ be the A-submodule of homomorphisms whose kernel contains P, or equivalently, such that $P\varphi = 0$. The module S is the socle of $\mathrm{Hom}_B(A, E)$ as an A-module. If $\varphi \in S$, then since $P_B A \subset P$, the image of φ is annihilated by P_B; that is, the image of φ is in the socle of E as a B-module, and since E is the injective hull of k_B, this is k_B. Since the homomorphisms in S all factor through the projection $A \to A/P = k$, we have $S \cong \mathrm{Hom}_B(k, k_B) \cong k$.

If $\varphi : A \to E$ is any B-module homomorphism, then since P is nilpotent, φ is annihilated by a power of P, and thus there is a multiple $a\varphi \neq 0$ that is annihilated by P. Thus S is an essential A-submodule of $\mathrm{Hom}_B(A, E)$, as required. □

21.2 Zero-Dimensional Gorenstein Rings

From the point of view of duality, the simplest case is the following:

Definition. *A zero-dimensional local ring A is* ***Gorenstein*** *if $A \cong \omega_A$.*

For example, fields are Gorenstein, and we shall soon see that complete intersections are Gorenstein.

The name *Gorenstein* comes from a duality property of Gorenstein rings that has as a special case a phenomenon concerning the singularities of plane curves that was studied by Daniel Gorenstein in his thesis [1952] under Oscar Zariski. The theory was cast in its current form through work of Serre and Bass (Gorenstein, who was later famous for his work on finite groups, always claimed that he couldn't even understand the definition of Gorenstein ring); aspects of it can be traced back to Macaulay's theory of

"inverse systems." Gorenstein rings are geometrically common and significant, as the title of Bass' foundational paper "On the Ubiquity of Gorenstein Rings" [1963] attests. There are many characterizations, among them the following in the zero-dimensional case:

Proposition 21.5. *Let A be a zero-dimensional local ring. The following are equivalent:*

a. A is Gorenstein.

b. A is injective as an A-module.

c. The socle of A is simple.

d. ω_A can be generated by one element.

Condition b is sometimes stated by saying that A is **self-injective**.

Proof.

a $\Rightarrow$ b: ω_A is injective as an A-module, so if $A \cong \omega_A$, then A is injective as an A-module.

b $\Rightarrow$ c: As A is a local ring, it is indecomposable as an A-module. If it is injective, it must be the injective hull of its socle. The injective hull of a direct sum is the direct sum of the injective hulls of the summands, so the socle must be simple.

c $\Rightarrow$ d: If the socle of A is simple, that is, isomorphic to A/P, then the "top" of the dual of A, that is the top of ω_A, which is $\omega_A/P\omega_A$, is simple. By Nakayama's lemma ω_A can be generated by one element.

d $\Rightarrow$ a: If ω_A is generated by one element then it is a homomorphic image of A. But A and ω_A have the same length by Proposition 21.2, so $A \cong \omega_A$. □

In the examples of zero-dimensional rings pictured earlier, we see that $k[x,y]/(x^2,y^3)$ is Gorenstein while $k[x,y]/(x^2,x^2y,y^3)$ is not. In codimension 2 all Gorenstein rings are complete intersections. We shall give a codimension 3 example that is not a complete intersection in Example 21.7.

Most of the common methods of constructing Gorenstein rings work just as well in the case where A is not zero-dimensional, and we shall postpone them for a moment. However, one technique, Macaulay's method of **inverse systems**, is principally of interest in the zero-dimensional case. Let $S = k[x_1,\ldots,x_r]$. For each $d \geq 0$ let S_d be the vector space of forms of degree d in the x_i. Let $T = k[x_1^{-1},\ldots,x_r^{-1}] \subset K(S) = k(x_1,\ldots,x_r)$ be the polynomial ring on the inverses of the x_i.

We make T into an S-module as follows: Let $L \subset K(S)$ be the vector space generated by the monomials in the x_i that are not in T. Notice that

L is an S-submodule of $K(S)$. We identify the S-module $K(S)/L$ with T by means of the maps $T \subset K(S) \twoheadrightarrow K(S)/L$. More directly put: If $m \in S$ and $n \in T$ are monomials, then $m \cdot n$ is the monomial $mn \in K(S)$ if this happens to lie in T, and 0 otherwise.

Theorem 21.6. *With notation as above, there is a one-to-one inclusion reversing correspondence between finitely generated S-submodules $M \subset T$ and ideals $I \subset S$ such that $I \subset (x_1, \ldots, x_r)$ and S/I is a local zero-dimensional ring, given by*

$$M \mapsto (0 :_S M), \text{ the annihilator of } M \text{ in } S;$$
$$I \mapsto (0 :_T I), \text{ the submodule of } T \text{ annihilated by } I.$$

If M and I correspond then $M \cong \omega_{S/I}$, so the ideals $I \subset (x_1, \ldots, x_r)$ such that S/I is a local zero-dimensional Gorenstein ring are precisely the ideals of the form $I = (0 :_S f)$ for some nonzero element $f \in T$.

The element f associated in this way to a zero-dimensional Gorenstein ring S/I is sometimes called the **dual socle generator** of S/I.

Proof. The S-module T may be identified with the graded dual $\oplus_d \operatorname{Hom}_k(S_d, k)$ of S; indeed, the dual basis vector to $m \in S_d$ is $m^{-1} \in T$. Exercise A3.4 shows that this graded dual is the injective envelope of $k = S/(x_1, \ldots, x_r)$ as an S-module. In the following we shall use only this property.

It follows by Proposition 21.4 that for any local zero-dimensional quotient $A = S/I$ of S we have $\omega_A := \operatorname{Hom}_S(A, T)$, which we may think of as the submodule of T consisting of elements annihilated by I. By Corollary 21.3 the annihilator of ω_A in A is 0. Thus, we see that $I = (0 :_S \omega_A)$, so that starting from an ideal I we get back to I under the correspondence.

Now suppose that M is any submodule of T, let $I = (0 :_S M)$, and set $A = S/I$. By what we have already done it suffices to show that $M = (0 :_T I)$. Note that the annihilator of M in A is 0 by construction. Clearly, we have $M \subset (0 :_T I) = \omega_A$. Dualizing the inclusion, we get $A \twoheadrightarrow D(M)$. The kernel of this map would annihilate $D(M)$, and thus M, so the kernel is zero and the map is an isomorphism. Dualizing again, we get $M = \omega_A$ as required. □

Example 21.7. In the case $r = 3$, let $f = x_1^{-1}x_2^{-1} + x_3^{-2}$. We have $I := (0 :_S f) = (x_1^2, x_2^2, x_1x_3, x_2x_3, x_1x_2 - x_3^2)$. It is easy to check that the socle of S/I is generated by x_1x_2 (which is congruent mod I to x_3^2). Note that $k[x_1, x_2, x_3]/I$ is a local Gorenstein ring that is not a complete intersection. See Exercise 21.6 for a more systematic investigation of this example.

For a different view of inverse systems, see Exercise 21.7.

21.3 Canonical Modules and Gorenstein Rings in Higher Dimension

We now extend the preceding zero-dimensional theory to local Cohen-Macaulay rings of any dimension. Many questions about Cohen-Macaulay rings can be reduced to the zero-dimensional case, so it seems reasonable to give a definition in terms of the reduction.

Definition. *Let A be a local Cohen-Macaulay ring. A finitely generated A-module ω_A is a* ***canonical module for A*** *if there is a nonzerodivisor $x \in A$ such that $\omega_A/x\omega_A$ is a canonical module for $A/(x)$. The ring A is* ***Gorenstein*** *if A is itself a canonical module; that is, A is Gorenstein if there is a nonzerodivisor $x \in A$ such that $A/(x)$ is Gorenstein.*

The induction implicit in this definition terminates because $\dim A/(x) = \dim A - 1$. We may easily unwind the induction, and say that ω_A is a canonical module if some maximal regular sequence $x_1, \ldots, x_d$ on A is also an ω_A-sequence, and $\omega_A/(x_1, \ldots, x_d)\omega_A$ is the injective hull of the residue class field of $A/(x_1, \ldots, x_d)$. Similarly, A is Gorenstein iff $A/(x_1, \ldots, x_d)$ is a zero-dimensional Gorenstein ring for some maximal regular sequence $x_1, \ldots, x_d$. By Nakayama's lemma and Proposition 21.4d, this is the case iff A has a canonical module generated by one element.

For a simple example, consider the case when A is a regular local ring. We claim that A has a canonical module, and in fact that $\omega_A = A$. When $\dim A = 0$ the result is obvious, since A is a field. For the general case we do induction on the dimension. If we choose x in the maximal ideal of A, but not its square, then x is a nonzerodivisor and $A/(x)$ is again a regular local ring, so A/xA is a canonical module for $A/(x)$ and A is a canonical module for A, by definition.

There are three problems with these notions. First, it is not at all obvious from the definitions that they are independent of the nonzerodivisor x that was chosen. Second, something called a canonical module should at least be unique, and uniqueness is not clear either. Our first goal is to show that this independence and uniqueness do hold.

The third problem is that it is not obvious that a canonical module should exist. Here we are not quite so lucky: There are local Cohen-Macaulay rings with no canonical module. However, our second goal will be to establish that canonical modules do exist for any Cohen-Macaulay rings that are homomorphic images of regular local rings (and a little more generally). This includes complete local rings and virtually all other rings of interest in algebraic geometry and number theory.

Let A be a ring and let M be an A-module. We say that a complex of injective A-modules

$$\mathcal{E} : E_0 \xrightarrow{\psi_1} E_1 \to \cdots \to E_n \to \cdots$$

is an **injective resolution** of M if $\ker \psi_1 = M$ and $\mathcal{E}$ is exact except at E_0. If $\mathcal{E}$ is an injective resolution, then we say that $\mathcal{E}$ is a **minimal injective resolution** iff each E_n is the injective hull of ker $(E_n \to E_{n+1})$.

As is shown in Corollary A3.11, every module has a minimal injective resolution, which is unique up to isomorphism. The **injective dimension** $\mathbf{id}_A M$ of M is the length of this resolution (which may be ∞).

The next theorem solves the problem of the independence of the definition of the canonical module of the regular sequence chosen.

Theorem 21.8. *Let A be a local Cohen-Macaulay ring of dimension d, and let W be a finitely generated A-module. W is a canonical module for A iff*

a. $\operatorname{depth} W = d$.

b. W is a module of finite injective dimension (necessarily equal to d).

c. $\operatorname{End}(W) = A$.

We postpone the proof to develop some preliminary results. In condition c we have written $\operatorname{End}(W)$ for the $\operatorname{Hom}_A(W, W)$, the endomorphism ring of W.

As we shall see, in the presence of condition a, condition b means that W reduces modulo a system of parameters $x_1, \ldots, x_d$ to an injective module over $\bar{A} := A/(x_1, \ldots, x_d)$. Such a module is isomorphic to a direct sum of copies of $\omega_{\bar{A}}$. Condition c serves to limit the number of copies to 1. Thus condition c is a sort of rank-1 condition. In Exercise 21.18 it is shown that under good circumstances the canonical module is isomorphic to an ideal of A.

21.4 Maximal Cohen-Macaulay Modules

First, to clarify the meaning of condition a, we prove:

Proposition–Definition 21.9.. *Let A be a local ring of dimension d, and let M be a finitely generated A-module. The following conditions are equivalent:*

a. Every system of parameters in A is an M-sequence.

b. Some system of parameters in A is an M-sequence.

c. $\operatorname{depth} M = d$.

If these conditions are satisfied, we say that M is a **maximal Cohen-Macaulay module over A**. *Every element outside the minimal primes of A is a nonzerodivisor on M.*

Proof. The implications a $\Rightarrow$ b $\Rightarrow$ c are immediate from the definitions. Suppose depth $M = d$. If $x_1, \ldots, x_d$ is a system of parameters then $(x_1, \ldots, x_d)$ contains a power of the maximal ideal P of A. By Corollary 17.8, $\operatorname{depth}((x_1, \ldots, x_d), M) = \operatorname{depth}(P, M) = d$, so $x_1, \ldots, x_d$ is a regular sequence on M by Corollary 17.7.

To prove the last statement, note that if x_1 is not in any minimal prime of A, then $\dim A/(x_1) = \dim A - 1$, so a system of parameters mod x_1 may be lifted to a system of parameters for A beginning with x_1. Thus x_1 is a nonzerodivisor on M. □

In case A is zero-dimensional, all finitely generated modules are maximal Cohen-Macaulay modules. On the other hand, if A is a regular local ring, then by the Auslander-Buchsbaum formula the maximal Cohen-Macaulay A-modules are exactly the free A-modules.

More generally, if A is a finitely generated module over some regular local ring S of dimension d, then by the Auslander-Buchsbaum theorem the maximal Cohen-Macaulay modules over A are those A-modules that are free as S-modules. Thus maximal Cohen-Macaulay modules may be thought of as representations of A as a ring of matrices over a regular local ring—as such they generalize the objects studied in integral representation theory of finite groups under the name of **lattices**. We shall later exploit the following example: If $B = A/J$ is a homomorphic image of A such that B is again Cohen-Macaulay of dimension d as a ring, then B is a Cohen-Macaulay A-module.

21.5 Modules of Finite Injective Dimension

We next study condition b of Theorem 21.8. If E is an injective module and $J \subset A$ is an ideal, then by Lemma A3.8 the module $\operatorname{Hom}_A(A/J, E) = (0 :_E J) \subset E$ is an injective module over A/J.

Proposition 21.10. *Suppose $\mathcal{E}$ as in Section 21.3 is a minimal injective resolution of a module M, and that an element $x \in A$ is a nonzerodivisor on A and on M. Set $E_i' = \operatorname{Hom}_A(A/x, E_i)$. The complex*

$$\mathcal{E}' : E_1' \to E_2' \to \cdots,$$

whose maps are induced by those of $\mathcal{E}$, is a minimal injective resolution of M/xM over $A/(x)$. Thus $\operatorname{id}_{A/(x)} M/xM = \operatorname{id}_A M - 1$, and if N is an A-module annihilated by x, then

$$\operatorname{Ext}^j_{A/x}(N, M/xM) \cong \operatorname{Ext}^{j+1}_A(N, M) \quad \textit{for all } j \geq 0.$$

Proof. The homology of the complex

$$\operatorname{Hom}_A(A/x, \mathcal{E}) : E'_0 \to E'_1 \to E'_2 \to \cdots$$

is by definition $\operatorname{Ext}^*_A(A/x, M)$. On the other hand, M is an essential submodule of E_0, and M contains no submodule annihilated by x, so E_0 contains no submodule annihilated by x. Thus $E'_0 = 0$, and we see that $\operatorname{Hom}_A(A/x, \mathcal{E}) = \mathcal{E}'$.

Computing $\operatorname{Ext}^*_A(A/x, M)$ instead from the free resolution

$$0 \to A \xrightarrow{x} A \to A/(x) \to 0,$$

we see that $\operatorname{Ext}^1_A(A/(x), M) = M/xM$, while $\operatorname{Ext}^j_A(A/(x), M) = 0$ for $j \neq 1$. Thus $\mathcal{E}'$ is an injective resolution of M/xM. Note that the numbering of the terms of $\mathcal{E}'$ is such that $\operatorname{Ext}^j_{A/(x)}(N, M/xM)$ is the homology of $\operatorname{Hom}_{A/(x)}(N, \mathcal{E}')$ at $\operatorname{Hom}_{A/(x)}(N, E'_{j+1})$; strictly speaking we should say that $\mathcal{E}'[1]$ is an injective resolution of M/xM.

To see that $\mathcal{E}'$ is minimal, note that the kernel of $E'_n \to E'_{n+1}$ is the intersection of the essential submodule $\ker E_n \to E_{n+1}$ with E'_n, and is thus essential in E'_n.

It follows at once that $\operatorname{id}_{A/(x)} M/xM = \operatorname{id}_A M - 1$. If x annihilates the A-module N, then every map from N to an E_i has image killed by x, so $\operatorname{Hom}_A(N, \mathcal{E}) = \operatorname{Hom}_A(N, \mathcal{E}') = \operatorname{Hom}_{A/(x)}(N, \mathcal{E}')$. Taking homology, and taking into account the shift in numbering, we get the last statement of the proposition. □

To exploit this result, we need to know the modules of finite injective dimension over a zero-dimensional ring.

Proposition 21.11. *Let A be a local Cohen-Macaulay ring. If M is a maximal Cohen-Macaulay module of finite injective dimension, then* $\operatorname{id}_A M = \dim A$. *If* $\dim A = 0$, *then M is a direct sum of copies of ω_A, and $M \cong \omega_A$ iff* $\operatorname{End}_A(M) = A$.

Proof. Suppose first that $\dim A = 0$. Let $D = \operatorname{Hom}_A(-, \omega_A)$ be the dualizing functor. If M has finite injective dimension, then applying D to an injective resolution of M we see that $D(M)$ is a module of finite projective dimension, and is thus free by the Auslander-Buchsbaum formula (Theorem 19.9). Applying D again we see that $M = D^2M$ is a direct sum of copies of $DA = \omega_A$.

Using D, we see that the endomorphism ring of ω_A^n is the same as the endomorphism ring of A^n. Thus it is equal to A iff $n = 1$.

If $\dim A = d$ is arbitrary, then we may choose a regular sequence $x_1, \ldots, x_d$ of A that is a regular sequence on M, and use Proposition 21.10 d times to conclude that $\operatorname{id}_A M = d + \operatorname{id}_{A/(x_1,\ldots,x_d)} M/(x_1, \ldots, x_d)M = d + 0 = d$. □

Finally, to understand condition c of Theorem 21.8 (and for a uniqueness result), we need:

Proposition 21.12. *Let A be a local Cohen-Macaulay ring of dimension d, and let M be a maximal Cohen-Macaulay module of finite injective dimension.*

a. *If N is a finitely generated module of depth e, then $\operatorname{Ext}_A^j(N, M) = 0$ for $j > d - e$.*

b. *If x is a nonzerodivisor on M, then x is a nonzerodivisor on $\operatorname{Hom}_A(N, M)$. If N is also a maximal Cohen-Macaulay module, then*

$$\operatorname{Hom}_A(N, M)/x \operatorname{Hom}_A(N, M) \cong \operatorname{Hom}_{A/(x)}(N/xN, M/xM)$$

by the homomorphism taking the class of a map $\varphi : N \to M$ to the map $N/xN \to M/xM$ induced by φ.

Proof.

a. We do induction on e. By Proposition 21.11, the injective dimension of M is d, so that $\operatorname{Ext}_A^j(N, M) = 0$ for any N if $j > d$. This gives the case $e = 0$.

Now suppose $e > 0$, and let x be a nonzerodivisor on N that lies in the maximal ideal of A. From the short exact sequence

$$0 \to N \xrightarrow{x} N \to N/xN \to 0$$

we get a long exact sequence

$$\cdots \to \operatorname{Ext}_A^j(N, M) \xrightarrow{x} \operatorname{Ext}_A^j(N, M) \to \operatorname{Ext}_A^{j+1}(N/xN, M) \to \cdots .$$

The module N/xN has depth $e-1$, so by induction $\operatorname{Ext}_A^{j+1}(N/xN, M)$ vanishes if $j + 1 > d - (e - 1)$, that is, if $j > d - e$. By Nakayama's lemma, $\operatorname{Ext}_A^j(N, M)$ vanishes if $j > d - e$.

b. Suppose x is a nonzerodivisor on M. From the short exact sequence

$$0 \to M \xrightarrow{x} M \to M/xM \to 0$$

we derive a long exact sequence beginning

$$0 \to \operatorname{Hom}_A(N, M) \xrightarrow{x} \operatorname{Hom}_A(N, M) \to \operatorname{Hom}_A(N, M/xM) \to \operatorname{Ext}_A^1(N, M) \to \cdots .$$

Thus x is a nonzerodivisor on $\operatorname{Hom}_A(N, M)$. If N is a maximal Cohen-Macaulay module then $\operatorname{depth}(N) = d$, so $\operatorname{Ext}_A^1(N, M) = 0$ by part a. Every homomorphism $N \to M/xM$ factors uniquely through N/xN, so $\operatorname{Hom}_A(N, M/xM) = \operatorname{Hom}_A(N/xN, M/xM)$. The short exact sequence above thus becomes

$$0 \to \operatorname{Hom}_A(N, M) \xrightarrow{x} \operatorname{Hom}_A(N, M) \to \operatorname{Hom}_A(N/xN, M/xM) \to 0.$$

Since $\operatorname{Hom}_A(M/xM, N/xN) = \operatorname{Hom}_{A/(x)}(M/xM, N/xN)$, this proves part b. □

Proposition 21.13. *Let A be a local ring, and let M and N be finitely generated modules. Suppose that x is a nonzerodivisor on M and that x is in the maximal ideal of A. If $\varphi : N \to M$ is a map and $\psi : N/xN \to M/xM$ is the map induced by φ, then:*

a. If ψ is an epimorphism, then φ is an epimorphism.

b. If ψ is a monomorphism, then φ is a monomorphism.

Further, if M and N are maximal Cohen-Macaulay modules, M has finite injective dimension, and $\psi : N/xN \to M/xM$ is any map, then there is a map $\varphi : N \to M$ inducing ψ.

Proof.

a. If ψ is an epimorphism, then φ is an epimorphism by Nakayama's lemma.

b. Suppose that ψ is a monomorphism, and let $J = \ker \varphi$. Since J goes to zero in N/xN, we must have $J \subset xN$. On the other hand, since x is a nonzerodivisor on the image of φ, we have $(J :_N x) = J$. Thus $xJ = J$, so $J = 0$ by Nakayama's lemma.

The last statement is a restatement of part b of Proposition 21.12. □

Proof of Theorem 21.8. First suppose that W is a canonical module. We do induction on the dimension of A.

If $\dim A = 0$, then condition a is vacuous, condition b is satisfied because $W = \omega_A$ is injective, and condition c follows because, by duality $\operatorname{End}(\omega_A) = \operatorname{End}(D(\omega_A)) = \operatorname{End}(A) = A$.

Now suppose $\dim A > 0$, and let x be a nonzerodivisor. By hypothesis, W/xW is a canonical module over $A/(x)$, and by induction it satisfies conditions a, b, and c as an $A/(x)$-module.

Since x is a nonzerodivisor on W and W/xW has depth $d-1$, condition a is satisfied. W has finite injective dimension by Proposition 21.10.

Let $B = \operatorname{End}(W)$, and consider the natural map $\varphi : A \to B$ sending each element $a \in A$ to the map "multiplication by a" $\in \operatorname{End}(W)$. We must show that φ is an isomorphism. By Proposition 21.12, x is a nonzerodivisor on B, and $B/xB = \operatorname{End}(W/xW) = A/(x)$. Thus by induction the map φ induces an isomorphism $A/(x) \to B/xB$. It follows by Proposition 21.13 that φ is an isomorphism.

Next suppose that W is an A-module satisfying conditions a, b, and c. Again we do induction on $\dim A$.

In case $\dim A = 0$ we must show that $W = \omega_A$. By Proposition 21.11 this follows from conditions b and c.

Now suppose that $\dim A > 0$, and let x be a nonzerodivisor in A. The element x is also a nonzerodivisor on W by Proposition 21.9, so W/xW has depth $d-1$ over $A/(x)$. By Proposition 21.10, $\mathrm{id}_{A/(x)} W/xW < \infty$, and by Proposition 21.12, $\mathrm{End}(W/xW) = \mathrm{End}(W)/x\,\mathrm{End}(W) = A/(x)$. Thus W/xW is a canonical module for $A/(x)$ by induction, and W is a canonical module for A. □

21.6 Uniqueness and (Often) Existence

These results imply a strong uniqueness result.

Corollary 21.14 (Uniqueness of canonical modules). *Let A be a local Cohen-Macaulay ring with a canonical module W. If M is any finitely generated maximal Cohen-Macaulay A-module of finite injective dimension, then M is a direct sum of copies of W. In particular, any two canonical modules of A are isomorphic.*

Proof. We do induction on $\dim A$, the case $\dim A = 0$ being Proposition 21.11. If $x \in A$ is a nonzerodivisor, then x is a nonzerodivisor on W and on M, and $M/xM \cong (W/xW)^n$ for some n. By Proposition 21.13, there is an isomorphism $M \cong W^n$. □

Henceforth we shall write ω_A for a canonical module of A (if one exists).

We now come to the question of existence. We have already seen that if R is a regular local ring, then R has canonical module $\omega_R = R$. We shall now show that if A is a homomorphic image of a local ring with a canonical module, then A has a canonical module too.

Theorem 21.15 (Construction of canonical modules). *Let R be a local Cohen-Macaulay ring with canonical module ω_R. If A is a local R-algebra that is finitely generated as an R-module, and A is Cohen-Macaulay, then A has a canonical module. In fact, if $c = \dim R - \dim A$, then*

$$\omega_A \cong \mathrm{Ext}^c_R(A, \omega_R)$$

is a canonical module for A.

Proof. We shall do induction on $\dim A$. First suppose that $\dim A = 0$. In this case c is the dimension of R. The annihilator of A contains a power of the maximal ideal of R. By Corollary 17.8 we may choose a regular sequence $x_1, \ldots, x_c$ of length c in the annihilator of A. Let $R' = R/(x_1, \ldots, x_c)$. R' is a local Cohen-Macaulay ring of dimension 0, and A is a finitely generated R'-module.

By definition, $\omega_R/(x_1, \ldots, x_c)\omega_R$ is a canonical module for R', for which we shall write $\omega_{R'}$. By Proposition 21.10, applied c times,

$$\operatorname{Ext}^c_R(A,\omega_R) \cong \operatorname{Ext}^0_{R'}(A,\omega_{R'}) = \operatorname{Hom}_{R'}(A,\omega_{R'}).$$

By Proposition 21.4, this is a canonical module for A, as required.

Now suppose that $\dim A > 0$. It suffices to show that if x is a nonzerodivisor on A, then x is a nonzerodivisor on $\operatorname{Ext}^c_R(A,\omega_R)$ and $\operatorname{Ext}^c_R(A,\omega_R)/x\operatorname{Ext}^c_R(A,\omega_R)$ is a canonical module for A/x. The short exact sequence

$$0 \to A \xrightarrow{x} A \to A/(x) \to 0$$

gives rise to a long exact sequence in Ext of which a part is

$$\begin{aligned} \cdots \to \ & \operatorname{Ext}^c_R(A/(x),\omega_R) \to \operatorname{Ext}^c_R(A,\omega_R) \xrightarrow{x} \operatorname{Ext}^c_R(A,\omega_R) \\ \to \ & \operatorname{Ext}^{c+1}_R(A/(x),\omega_R) \to \operatorname{Ext}^{c+1}_R(A,\omega_R) \to \cdots. \end{aligned}$$

By induction, $\operatorname{Ext}^{c+1}_R(A/(x),\omega_R)$ is a canonical module for $A/(x)$, so it suffices to show that the two outer terms are 0, which we may do as follows:

Set $I = \operatorname{ann}_R A$. The ring $A/(x)$ is annihilated by (I,x), which has depth $c+1$ in R. Thus $\operatorname{Ext}^c_R(A/(x),\omega_R) = 0$ by Proposition 18.4.

The ring A, being Cohen-Macaulay, has depth equal to $\dim R - c$ so $\operatorname{Ext}^{c+1}_R(A,\omega_R) = 0$ by Proposition 21.12a. □

This description of ω_A has a particularly simple interpretation when R is a regular local ring:

Corollary 21.16. *Let R be a regular local ring, suppose that I is an ideal of codimension c in R, and suppose that $A = R/I$ is Cohen-Macaulay. If $\mathcal{F}$ is the minimal free resolution of A as an R-module, then the length of $\mathcal{F}$ is c, and $\mathcal{F}^*$ is the minimal free resolution of ω_A. If we write $\mathcal{F}$ as*

$$\mathcal{F} : 0 \to F_n \xrightarrow{\varphi_c} F_{n-1} \to \cdots \to F_1 \xrightarrow{\varphi_1} R,$$

then $n = c$ and the following statements are equivalent:

a. A is Gorenstein.

b. $\mathcal{F}$ is symmetric in the sense that $\mathcal{F}^* \cong \mathcal{F}$ as complexes.

c. $F_c \cong R$.

Proof. By Corollary 19.15, $\operatorname{pd}_R A = c$. By Proposition 18.4 $\operatorname{Ext}^j_R(A,R) = 0$ for $j < c$, so $\mathcal{F}^*$ is a minimal free resolution of $\operatorname{Ext}^c_R(A,R) = \omega_A$.

a ⇒ b: If A is Gorenstein then $\omega_A \cong A$, so $\mathcal{F}^*$ is another minimal free resolution of A, and part b follows from the uniqueness of minimal free resolutions.

b ⇒ c: By b, $F_c = F_0^* = R^* = R$.

c ⇒ a: From c we see that ω_A can be generated by one element so $\omega_A = A/J$ for some ideal J. Since End $(\omega_A) = A$ by Theorem 21.8, $J = 0$. (Alternatively, we could reduce to the zero-dimensional case and use Proposition 21.5d.) □

21.7 Localization and Completion of the Canonical Module

Corollary 21.17. *Let A be a local Cohen-Macaulay ring with canonical module ω_A. If P is a prime ideal of A, then $(\omega_A)_P$ is a canonical module of A_P; in particular, if A is Gorenstein, then A_P is Gorenstein.*

Proof. We prove that each of the properties in the characterization of Theorem 21.8 localizes:

a. Set $\operatorname{codim} P = c$. We claim that there is a system of parameters for A whose first c elements lie in P. It suffices to show that some sequence of elements $x_1, \ldots, x_c \in P$ generates an ideal of height c. Let $Q \subset P$ be a prime of codimension $c - 1$. By induction we may find $x_1, \ldots, x_{c-1} \in Q$ generating an ideal I of codimension $c - 1$. The principal ideal theorem (Theorem 10.2) shows that each of the finitely many primes minimal over I has codimension $\leq c - 1$, and therefore does not contain P. Using prime avoidance (Lemma 3.3) we see that P is not in the union of these primes, so we can choose $x_c \in P$ not in any of the minimal primes of I. It follows that $x_1, \ldots, x_c$ generates an ideal of codimension c.

   Now suppose that M is any finitely generated module of depth d—that is, a maximal Cohen-Macaulay module over A. It follows from Proposition 21.9 that $x_1, \ldots, x_c$ is an M-sequence. Localizing at P preserves this property, so we see that M_P is again a maximal Cohen-Macaulay module over A_P. This shows that $(\omega_A)_P$ satisfies property a over A_P.

b. We claim that if we localize a finite injective resolution of ω_A we get an injective resolution over A_P of ω_P. It suffices to show that if E is an injective A-module, then E_P is an injective A_P-module. Every ideal of A_P has the form I_P for some ideal I of A. If $\varphi : I_P \to E_P$ is a homomorphism, we must show that φ extends to a map $A_P \to E_P$. Since I is finitely presented, Proposition 2.10 shows that $\operatorname{Hom}_{A_P}(I_P, E_P) = \operatorname{Hom}_A(I, E)_P$. Thus there is an element $u \notin P$ such that $u\varphi$ is the localization of some map $\psi : I \to E$. Since E is injective, ψ extends to $\psi' : A \to E$, and $u^{-1}\psi'$ is the desired extension of φ.

c. Since ω_P is a finitely presented module, $\operatorname{End}_{A_P}((\omega_A)_P) = \operatorname{End}_A(\omega_A)_P = A_P$ by Proposition 2.10. □

Corollary 21.18. *If A is a local Cohen-Macaulay ring with canonical module ω_A, then $\omega_{\hat{A}} = (\omega_A)^\wedge$, the completion of ω_A. In particular, A is Gorenstein iff the completion $\hat{A}$ is Gorenstein.*

Proof. It is enough, by the definition of the canonical module, to show that for some system of parameters $x_1, \dots, x_d$ the module $(\omega_A)^\wedge/(x_1, \dots, x_d)(\omega_A)^\wedge$ is a canonical module for $\hat{A}/(x_1, \dots, x_d)\hat{A}$. If we choose $x_1, \dots, x_d$ in A, then $A/(x_1, \dots, x_d)$, being Artinian, is already complete, so $\hat{A}/(x_1, \dots, x_d)\hat{A} = A/(x_1, \dots, x_d)$. The same goes for any module over $A/(x_1, \dots, x_d)$, so $(\omega_A)^\wedge/(x_1, \dots, x_d)(\omega_A)^\wedge = \omega_A/(x_1, \dots, x_d)\omega_A$, which is the canonical module of $A/(x_1, \dots, x_d)$ as required.

Since $\hat{A}/(x_1, \dots, x_d)\hat{A} = A/(x_1, \dots, x_d)$, one is Gorenstein iff the other is. The second statement follows.

21.8 Complete Intersections and Other Gorenstein Rings

We next turn to some examples. The most common examples of Gorenstein rings are complete intersections—regular local rings modulo regular sequences.

Corollary 21.19. *If $A = R/I$ where R is a regular local ring and I is an ideal generated by a regular sequence, then A is Gorenstein.*

Proof. If $x_1, \dots, x_c$ is a regular sequence generating I, then the Koszul complex $K(x_1, \dots, x_c)$ is the minimal free resolution of A as an R-module, so $\operatorname{pd}_R A = c, \operatorname{Ext}^c_R(A, R) \cong A$, and we are done by Theorem 21.15.

The converse is false: We have already seen an example of a Gorenstein ring $A = R/I$ where I had codimension 3 and was generated by five quadrics. However, there is no such example in codimension 2.

Corollary 21.20 (Serre). *Let $A = R/I$ where R is a regular local ring. If* $\operatorname{codim} I = 1$*, then A is Cohen-Macaulay iff A is Gorenstein iff I is principal. If* $\operatorname{codim} I = 2$*, then A is Gorenstein iff I is generated by a regular sequence of length* 2.

Proof. Let $K(R)$ be the quotient field of R. Since I is nonzero, $K(R) \otimes R/I = 0$, so that with notation as in Corollary 21.16, $K(R) \otimes \mathcal{F}$ is an exact sequence of vector spaces, and we see that the alternating sum of the ranks of the free modules in $\mathcal{F}$ is 0.

Suppose $\operatorname{codim} I = 1$. If A is Cohen-Macaulay, then $\operatorname{pd}_R A = 1$ and $\mathcal{F}$ has the form

$$0 \to F_1 \to R \to A \to 0.$$

Thus rank $F_1 = 1$—that is, I is principal. Conversely, if I is principal, then A is Gorenstein by Corollary 21.19.

If $\operatorname{codim} I = 2$, then the resolution has the form

$$0 \to F_2 \to F_1 \to R \to R/I \to 0.$$

If A is Gorenstein, then F_2 has rank 1, and we deduce that F_1 has rank 2. Thus I is generated by two elements that must form a regular sequence by Corollary 17.7. Again, the converse is given by Corollary 21.19.

Even in higher codimension the ideals I in a regular local ring R such that R/I is Gorenstein are subject to some surprising restrictions: For example, if $\operatorname{codim} I = c$ then of course I may be minimally generated by c elements (a regular sequence) but not by $c+1$ elements; see Kunz [1974]. Also, if $c = 3$, the minimal number of generators must be odd, and in fact there is a structure theorem along the lines of the Hilbert-Burch theorem (Buchsbaum and Eisenbud [1977]). In codimension $c \geq 4$ however, every minimal number of generators $\geq c+2$ is possible; the restrictions are more subtle and are still a matter of current research.

21.9 Duality for Maximal Cohen-Macaulay Modules

The duality theory for modules of finite length generalizes to a duality for maximal Cohen-Macaulay modules. We shall describe this generalization now to prepare for the idea of linkage explained in the next section.

Theorem 21.21 (Duality). *Let A be a local Cohen-Macaulay ring and let D be the functor $\operatorname{Hom}_A(-, \omega_A)$. The functor D is a dualizing functor on the category of maximal Cohen-Macaulay A-modules in the sense that*

a. *D takes maximal Cohen-Macaulay A-modules to maximal Cohen-Macaulay A-modules.*

b. *D takes exact sequences of maximal Cohen-Macaulay A-modules to exact sequences.*

c. *The natural map $M \to D^2M = \operatorname{Hom}_A(\operatorname{Hom}_A(M, \omega_A), \omega_A)$ sending $m \in M$ to the map $\alpha \mapsto \alpha(m)$ for $\alpha \in \operatorname{Hom}_A(M, \omega_A)$ is an isomorphism when M is a maximal Cohen-Macaulay A-module.*

Proof.

a. This follows by induction on dim A, using Proposition 21.12b.

b. If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of A-modules, then we get an exact sequence

$$\begin{aligned} 0 \to \operatorname{Hom}_A(M'', \omega_A) &\to \operatorname{Hom}_A(M, \omega_A) \to \operatorname{Hom}_A(M', \omega_A) \\ \to \operatorname{Ext}^1_A(M'', \omega_A) &\to \cdots. \end{aligned}$$

If in addition M'' is a maximal Cohen-Macaulay module, then $\operatorname{Ext}^1_A(M'', \omega_A) = 0$ by Proposition 21.12a.

c. Let φ_M be the natural map $M \to D^2(M) = \mathrm{Hom}_A(\mathrm{Hom}_A(M, \omega_A), \omega_A)$ sending $m \in M$ to the map $\alpha \mapsto \alpha(m)$ for $\alpha \in \mathrm{Hom}_A(M, \omega_A)$. We shall prove by induction on $\dim A$ that φ_M is an isomorphism for every maximal Cohen-Macaulay module M. The case $\dim A = 0$ was treated in Proposition 21.2.

By Proposition 21.12, $D(M)/xD(M) \cong D'(M/xM)$, where $D' = \mathrm{Hom}_{A/x}(M/xM, \omega_A/x\omega_A)$ is the duality for maximal Cohen-Macaulay modules over $A/(x)$. Unraveling the isomorphism, we see that the map $M/xM \to D^2(M)/xD^2(M)$ induced by φ_M is the corresponding duality map $\varphi_{M/xM}$ defined over $A/(x)$. Since this map is an isomorphism by induction, it follows from Proposition 21.13a and b that φ_M is an isomorphism. □

21.10 Linkage

A striking application of Theorem 21.15 occurs in the theory of linkage as formulated by Peskine and Szpiro [1974]. The classical cases involve sets of points in the plane or curves in $\mathbf{P}^3$. The case of points in the plane is the setting of the Cayley-Bacharach theorem and its generalizations; we leave this to Exercise 21.24. We begin with an example from the case of curves. For simplicity we work over an algebraically closed field.

Suppose that X, $Y \subset \mathbf{P}^3$ are curves without common components. We say that X and Y are **directly linked** if $X \cup Y$ is a complete intersection in the sense that the ideal $I(X \cup Y)$ is generated by a regular sequence. We say that curves X and Z are **linked** if they can be connected by a sequence of such direct linkages. For example, the twisted cubic curve X, which is the image of $\mathbf{P}^1$ in $\mathbf{P}^3$ under the map

$$(s,t) \mapsto (s^3, s^2t, st^2, t^3),$$

has ideal generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x_0 & x_1 \\ x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}.$$

If one takes just two of these minors—say the two containing the first row in common, $f_1 = x_0x_2 - x_1^2$, $f_2 = x_0x_3 - x_1x_2$—it turns out that they generate the ideal of the union of X with one of its secant lines Y (Figure 21.1), in this case the line $x_0 = x_1 = 0$.

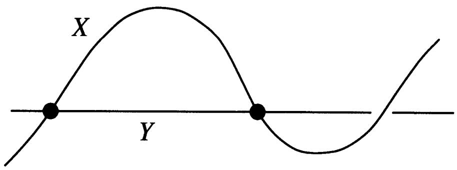


FIGURE 21.1. The twisted cubic curve X and one of its secants Y.

(Note that x_0, x_1 are the two 1×1 minors—that is, entries—of the common row. This is no accident, as we shall see in Proposition 21.24.) Thus X and Y are directly linked.

This sort of relationship between curves has been of interest for a long time, as evidenced by the article by Rohn and Berzolari in the German *Enzykolpädie der Mathematischen Wissenschaften* [1921–1928]. However, it was given new interest by the discoveries of Apéry [1945a, b] and Gaeta [1952], who showed essentially that if X and Z are curves in $\mathbf{P}^3$ whose homogeneous coordinate rings are Cohen-Macaulay, then X and Z are linked. Peskine and Szpiro [1974] greatly generalized and recast the theory in the form we give it below. Hartshorne and Rao (see Rao [1979]) then generalized the result of Apéry and Gaeta. They showed that the linkage classes of curves in $\mathbf{P}^3$ are distinguished by a simple invariant: If we write $S(X)$ for the homogeneous coordinate ring of the curve $X \subset \mathbf{P}^3$, then $S(X)$ is a homomorphic image of $S = k[x_0, \ldots, x_3]$, the coordinate ring of $\mathbf{P}^3$ itself, and two curves X and Z are linked iff the graded modules of finite length $\operatorname{Ext}^3_S(S(X), S)$ and $\operatorname{Ext}^3_S(S(Z), S)$ are equal up to a shift in grading and a dualization. (The theorem of Apéry and Gaeta is a special case, since $S(X)$ is Cohen-Macaulay iff $\operatorname{Ext}^3_S(S(X), S) = 0$.) See Martin-Deschamps and Perrin [1990] for a treatment of space curves that grows from this fact.

It is not hard to show that if X and Y are directly linked, and the ideal of the complete intersection $X \cup Y$ is (f_1, f_2), then $((f_1, f_2) : I(X)) = I(Y)$. In the treatment of Peskine and Szpiro, this algebraic relationship takes center stage. In general, they define ideals I and J of codimension c in a Gorenstein local (or suitably graded) ring A to be **directly linked** if there is a regular sequence $f_1, \ldots, f_c$ in A such that $J = ((f_1, \ldots, f_c) : I)$ and $I = ((f_1, \ldots, f_c) : J)$. I and J are **linked** if they are connected by a series of direct linkages.[2] We often say that A/I and A/J are linked to mean that I and J are.

Linkage is in some ways a rather crude equivalence relation. For example, we have:

[2]A priori, if we apply this notion in the case where A is the polynomial ring in four variables and $c = 2$, we get a weaker relation than the linkage defined for pairs of reduced curves without the common component above; but Peskine and Szpiro prove that the two notions coincide, and we shall henceforth use only this algebraic version.

Proposition 21.22. *Let A be a Gorenstein local ring. If $x_1, \ldots, x_c$ and $y_1, \ldots, y_c$ are regular sequences in A, then $(x_1, \ldots, x_c)$ and $(y_1, \ldots, y_c)$ are linked.*

Proof. We do induction on c, beginning with $c = 1$. If x and y are nonzero-divisors in A and $xa = xyb$, then we can cancel the x and get $a = yb$. Thus $(xy : x) = (y)$ and similarly $(xy : y) = (x)$, so we see that (x) and (y) are linked.

For arbitrary c, we first observe that if Q is an associated prime of $A/(y_1, \ldots, y_{c-1})$ then Q has depth $c - 1$. Since A is Cohen-Macaulay, Q has codimension $c - 1$ by Theorem 18.7. On the other hand, the ideal $(x_1, \ldots, x_c)$ has codimension c by Proposition 18.2. Thus $(x_1, \ldots, x_c)$ is not contained in any of the associated primes of $A/(y_1, \ldots, y_{c-1})$, and by the refined prime avoidance lemma of Exercise 3.19b, there is an element of the form $x'_c := x_c + \sum_1^{c-1} a_i x_i$ that is not in any of these associated primes. Replacing x_c by x'_c, we may assume from the outset that $y_1, \ldots, y_{c-1}, x_c$ is also a regular sequence.

It now suffices to prove the proposition in the case where the two regular sequences share a common element. By the permutability of regular sequences, we may reindex the elements, and take this common element to be the first, say $x_1 = y_1$. Working modulo x_1 the inductive hypothesis shows that $(x_1, y_2, \ldots, y_c)$ and $(x_1, \ldots, x_c)$ are linked by a sequence of direct linkages each involving a regular sequence beginning with x_1. □

On the other hand, there is a strong connection between directly linked rings. Here is the key result.

Theorem 21.23 (Linkage). *Let A be a Gorenstein local ring, and let I be an ideal of codimension 0. Set $J = (0 :_A I)B = A/I$ and $J = (0 :_A I)$. We have $J \cong \mathrm{Hom}_A(B, A)$.*

a. *The ideal J has codimension 0 and no embedded components. If I has no embedded components, then $I = (0 :_A J)$ so I and J are linked.*

b. *If $B := A/I$ is a Cohen-Macaulay ring, then $C := A/J$ is a Cohen-Macaulay ring.*

c. *If $B := A/I$ is a Cohen-Macaulay ring, then $J = (0 :_A I) = \mathrm{Hom}_A(B, A)$ is a canonical module for B; in particular, B is Gorenstein iff J is a principal ideal of A.*

Proof.

a. The map $\mathrm{Hom}_A(B, A) \to A$ sending a homomorphism φ to $\varphi(1)$ is an isomorphism onto $(0 :_A I)$, proving the first statement. If P is a prime associated to $(0 :_A I)$, then it is still associated after we localize at P, so we may assume that P is a maximal ideal. For some $y \notin (0 :_A I)$ we have $Py \subset (0 :_A I)$, that is, $PyI = 0$. Since $yI \neq 0$, this implies

that P annihilates an element of A, and is thus associated to 0 in A. Since A is Cohen-Macaulay, P must be a minimal prime.

Quite generally and trivially, $I \subset (0 :_A J)$. If I has no embedded components, then to prove equality it is enough to do so locally at a minimal prime of A. Thus we may assume that A is a zero-dimensional Gorenstein ring.

Note that $\operatorname{Hom}_A(A/I, A) = (0 :_A I)$ since each homomorphism may be identified with the image of $1 \in A/I$, and the elements that can be the images of 1 are just the elements of $(0 :_A I)$. Thus

$$J = (0 :_A I) = \operatorname{Hom}_A(A/I, A) = \operatorname{Hom}_A(A/I, \omega_A) = D(A/I),$$

where D is the dualizing functor. Since D preserves length we have $\operatorname{length} J = \operatorname{length} A/I = \operatorname{length} A - \operatorname{length} I$. Repeating the argument, we see that $\operatorname{length}(0 :_A J) = \operatorname{length} A/J = \operatorname{length} A - \operatorname{length} J = \operatorname{length} I$, so $I = (0 : J)$.

b. Suppose that A/I is Cohen-Macaulay. Let D be the functor $\operatorname{Hom}_A(-, \omega_A) = \operatorname{Hom}_A(-, A)$, as in Theorem 21.21. Since B and A are maximal Cohen-Macaulay A-modules, the same is true of I by Corollary 18.6b. Just as we saw that $J = D(B)$ above, we have $C = D(I)$, so C is a maximal Cohen-Macaulay A-module, and is thus a Cohen-Macaulay ring.

c. The equation $J = D(B)$ shows that $J = \omega_B$. The last statement follows by Proposition 21.5d and Nakayama's lemma. □

To bring the discussion back to earth, we give a modern presentation of the idea of Apéry [1945a, b] and Gaeta [1952] that initially gave rise to this whole theory. It generalizes the example of the twisted cubic and secant line with which we started. It is generally applied in the graded case where A is a polynomial ring over a field.

Proposition 21.24. *Let A be a local Gorenstein ring, and let M be an $(n+1) \times n$ matrix over A. Write Δ_i for the $n \times n$ minor of M obtained by omitting the ith row, and let M' be the result of dropping the first two rows from M, as in the following picture. If Δ_1, Δ_2 is a regular sequence, then $B = A/I_n(M)$ and $C = A/I_{n-1}(M')$ are Cohen-Macaulay rings, linked in the Gorenstein ring $A/(\Delta_1, \Delta_2)$.*

$$M = \overbrace{\begin{pmatrix} M'' \\ \hline \\ M' \\ \\ \end{pmatrix}}^{n} \begin{matrix} \}\, 2 \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n-1 \end{matrix}$$

Proof. Write I for $I_n(M)$ and J for $I_{n-1}(M')$. The minors Δ_1, Δ_2 may be written (by the "Laplace expansion" of determinants) as linear combinations of the $(n-1) \times (n-1)$ minors of M', so that $\Delta_1, \Delta_2 \in I \cap J$, and both I and J have depth ≥ 2. From the Hilbert-Burch theorem (Theorem 20.15) we see that we have resolutions

$$0 \to A^n \xrightarrow{M} A^{n+1} \to A \to B \to 0$$

and

$$0 \to A^{n-1} \xrightarrow{M'^*} A^n \to A \to C \to 0.$$

From the Auslander-Buchsbaum formula it follows that B and C are both Cohen-Macaulay of codimension 2 in A.

We shall show that $I/(\Delta_1, \Delta_2) = \omega_C$. Since the annihilator of ω_C is J, this implies that

$$((\Delta_1, \Delta_2) :_A I) = J$$

as required.

Examining the resolutions above we see that M is a presentation matrix for I. Supposing that the first two basis vectors of A^n map to Δ_1, Δ_2, we can get a presentation matrix for $I/(\Delta_1, \Delta_2)$ by adding two new relations that map to these first two basis vectors. Thus we have an exact sequence of the form

$$A^2 \oplus A^n \xrightarrow{\tilde{M}} A^{n+1} \to I/(\Delta_1, \Delta_2) \to 0$$

where $\tilde{M}$ is given as a matrix by

$$\tilde{M} = \left(\begin{array}{cc|c} 1 & 0 & \\ 0 & 1 & \smash{\overbrace{\qquad M'' \qquad}^{n}} \\ \hline 0 & 0 & \\ \cdot & \cdot & M' \\ 0 & 0 & \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\0\end{array}}\right\} 2 \\ \left.\vphantom{\begin{array}{c}0\\ \cdot \\ 0\end{array}}\right\} n-1 \end{array}$$

By adding multiples of the first two columns to the other columns, which corresponds to a change of basis in $A^2 \oplus A^n, \tilde{M}$ can be put in the form

$$\left(\begin{array}{cc|cccc} 1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & 0 \\ \hline 0 & 0 & & & & \\ \cdot & \cdot & & M' & & \\ 0 & 0 & & & & \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\0\end{array}}\right\} 2 \\ \left.\vphantom{\begin{array}{c}0\\ \cdot \\ 0\end{array}}\right\} n-1 \end{array}$$

(with the last n columns marked n)

from which it is clear that the cokernel of $\tilde{M}$ is the same as the cokernel of M'. Examining the preceding resolutions, we see that the cokernel of M' is $\operatorname{Ext}_A^2(C, A)$, and by Theorem 21.15 this is ω_C, as required.

Corollary 21.25. *If A is a regular local ring, then any two Cohen-Macaulay factor rings of codimension 2 are linked.*

Proof. Let $I \subset A$ be an ideal of codimension 2 such that A/I is Cohen-Macaulay. Proposition 21.22 shows that any two complete intersections are linked to one another, so it is enough to show that I is linked to an ideal generated by two elements.

Let m be the minimal number of generators of I. Using the prime avoidance lemma we may find a minimal set of generators for I such that the first two elements generate an ideal of codimension 2: Simply choose the first element outside the union of the minimal primes of A, and the second element outside the minimal primes of the first element (these have codimension 1 by the principal ideal theorem).

By the Hilbert-Burch theorem, there is an $(m-1) \times m$ matrix M whose $(m-1) \times (m-1)$ minors are the chosen generators of I. If $m > 2$ we apply Proposition 21.24 (with $n = m-1$) and conclude that I can be linked to an ideal generated by just $m-1$ elements. We may repeat this argument until we get an ideal generated by 2 elements. $\square$

The statement of Corollary 21.25 fails in the case of codimension 3: The ideal $(x, y, z)^2 \subset k[x, y, z]$ is not linked to a complete intersection. See Huneke [1984] and the references there. There is a growing body of evidence that Cohen-Macaulay factor rings that are linked to complete intersections are particularly nice. See, for example, Huneke and Ulrich [1987].

21.11 Duality in the Graded Case

Now suppose that instead of being local, A is a positively graded Cohen-Macaulay ring $A = A_0 \oplus A_1 \oplus \cdots$ with $A_0 = k$ a field. As with all the results in this book, the duality theory developed in this chapter can be transposed to the graded case. However, there is one point that requires attention: The module ω_A will now be a graded module, and we must distinguish between a module and its shifts. (Recall that if M is a graded module, then the module $M(\delta)$ obtained from M by shifting δ steps is the graded module with the same homogeneous components as M but renumbered so that $M(\delta)_n = M_{\delta+n}$.) For example, in the zero-dimensional case, we insist that ω_A is the injective hull of the residue field k, where k is concentrated in degree 0; thus for a field $A = k$, it is still true that $\omega_k \cong k$. In order to make the fundamental result

$$\omega_A = \operatorname{Ext}_R^c(A, \omega_R)$$

of Theorem 21.15 true, however, we must change the definition of the canonical module. To see why, consider the simple case $R = k[x]$, where x is an indeterminate of degree 1, and $A = k[x]/(x^t)$ for some integer $t > 0$. The socle of A is generated by x^{t-1}. Thus A is the injective hull of $k(-t+1)$, not of k, so ω_A, the injective hull of k, is $A(t-1)$. The graded free resolution

$$0 \to R(-t) \to R \to A \to 0$$

shows that $\operatorname{Ext}_R^1(A, R) = A(t)$, so $\omega_A = \operatorname{Ext}_R^1(A, R(-1))$. This suggests that we should have $\omega_R = R(-1)$.

Definition. *Let A be a positively graded Cohen-Macaulay ring $A = A_0 \oplus A_1 \oplus \cdots$ with $A_0 = k$ a field. A finitely generated A-module W is a* ***graded canonical module for A*** *if there is a homogeneous nonzerodivisor $x \in A$ of some degree δ such that $(W/xW)(\delta)$ is a canonical module for $A/(x)$. The ring A is* ***Gorenstein*** *if $A(\delta)$ is a canonical module for some δ; that is, A is Gorenstein if there is a homogeneous nonzerodivisor $x \in A$ such that $A/(x)$ is Gorenstein.*

Using this definition it is easy to show, for example, that if $A = k[x_1, \ldots, x_r]$, the polynomial ring on r indeterminates of degree 1, then $\omega_A = A(-r)$; one uses induction, factoring out one indeterminate at a

time. It follows from the definition that if we factor out a regular sequence $f_1, \ldots, f_c$ of homogeneous elements of degrees $\delta_1, \ldots \delta_c$, then the canonical module of $B := A/(f_1, \ldots, f_c)$ is $\omega_B = B(\Sigma\delta_i - r)$. (See Exercise 21.16 for a more general case and Exercise 21.22 for the effect that this has on linkage.)

Theorem 21.8 also requires some modification in the graded case. The conditions as they stand characterize the canonical module only up to a shift. One convenient way to change the list of properties to get a characterization is to replace property c by the property

c′. $\mathrm{Ext}^d_A(k, W) \cong k$.

See Exercise 21.13. We made the choice we did in the text because from the condition $\mathrm{End}(W) = A$ it is obvious that the annihilator of W is 0; and also because the condition $\mathrm{End}(W) = A$ localizes.

21.12 Exercises

The Zero-Dimensional Case and Duality

Exercise 21.1:* Let A be the ring $k[x, y]/(x^2, xy, y^n)$. Show that the socle of A is a 2-dimensional graded vector space with generators in degrees 1 and $n - 1$. Draw diagrams for A and $D(A)$ in the style of the pictures at the beginning of this chapter.

Exercise 21.2:* Let A be a ring, and let D be a contravariant A-linear functor from the category of A-modules of finite length to itself. If $D^2 \cong 1$ as functors, show that D is exact.

Exercise 21.3: Let A be a finite-dimensional algebra over a field k. Not assuming that A is local, show that $\mathrm{Hom}_k(A, k)$ is the injective hull of the direct sum of the simple A-modules.

Exercise 21.4:* Here is a different generalization of the duality theory in the zero-dimensional case:

Let R be a regular local ring of dimension d. Show that the functor $\mathrm{Ext}^d_R(-, \omega_R)$ is a dualizing functor on the category of R-modules of finite length.

Exercise 21.5:* Prove that the dualizing functor described in Exercise 21.4 is, up to isomorphism, the only dualizing functor on the category of R-modules of finite length.

Exercise 21.6: Let f be any quadratic form in $k[x_1^{-1}, x_2^{-1}, x_3^{-1}]$, and consider the ideal I_f in $S = k[x_1, x_2, x_3]$ corresponding to f. Show that

1. f has rank 1 as a quadratic form iff I_f is generated by a regular sequence consisting of two linear forms and a cubic.

2. f has rank 2 as a quadratic form iff I_f is generated by a regular sequence consisting of a linear form and two quadrics.

3. f has rank 3 as a quadratic form iff I_f has 5 quadratic generators.

Exercise 21.7 (Inverse systems and differential operators): Here is a more astonishing form of the theory of inverse systems. Let k be a field of characteristic 0, and let $T = k[y_1, \ldots, y_r]$ be the polynomial ring. Recall that a **polynomial differential operator with constant coefficients** over k is an operator on T of the form

$$D = \sum_{i_1,\ldots,i_r} a_{i_1,\ldots,i_r} (\partial/\partial y_1)^{i_1} \cdot \cdots \cdot (\partial/\partial y_r)^{i_r},$$

with $a_{i_1,\ldots,i_r} \in k$. The **symbol** of D is the polynomial (in a new polynomial ring $S = k[x_1, \ldots, x_r]$) obtained by replacing each $\partial/\partial y_i$ by x_i:

$$\text{symbol}(D) := \sum_{i_1,\ldots,i_r} a_{i_1,\ldots,i_r} x_1^{i_1} \cdot \cdots \cdot x_r^{i_r} \in S.$$

Let $0 \neq f \in T$ be a polynomial. Prove that if I is the set of symbols in S of polynomial differential operators with constant coefficients D on T such that $Df = 0$, then I is an ideal of S contained in $(x_1, \ldots, x_r)$, and S/I is a zero-dimensional local Gorenstein ring. Conversely, prove that if $J \subset (x_1, \ldots, x_r)$ is an ideal such that S/J is a zero-dimensional Gorenstein ring, and J' is the set of differential operators with symbols in J, then there is a solution $f \in T$ to the differential equations

$$Dg = 0 \quad \text{for all } D \in J',$$

with the property that every solution g of these equations is obtained by applying some differential operator to f.

Exercise 21.8: Let A be a finite-dimensional algebra over a field k. Show that a map $A \to \text{Hom}_k(A, k)$, sending $1 \mapsto \varphi$, is an isomorphism iff φ generates $\text{Hom}_k(A, k)$ as an A-module.

Exercise 21.9: Let A, P be a zero-dimensional local ring, finite-dimensional as a vector space over a field $k \cong A/P$. Suppose that A is Gorenstein. Let $\varphi : A \to k$ be a linear functional on A. Show that the symmetric bilinear form $\langle a, b \rangle = \varphi(ab)$ is nonsingular iff φ is nonzero on the socle of A. Show that this form is an invariant of A, (as a k-algebra); for example, if $k = \mathbf{R}$ then the signature of this form is an invariant of

A. (When $F = (f_1, \ldots, f_r) : \mathbf{R}^r \to \mathbf{R}^r$ is an analytic mapping taking 0 to 0, and $f_1, \ldots, f_r$ is a regular sequence in the ring of real analytic germs of functions at 0, then this signature is the local degree at the origin of F; that is, the number of times F wraps a small sphere about the origin in the source space around a small sphere about the origin in the target space. See Eisenbud and Levine [1977].)

Exercise 21.10:* Here is a generalization of Exercise 21.9:

Suppose that (A, P) is a zero-dimensional local ring that happens to be a finite-dimensional vector space over a field k. Proposition 21.1 shows that $\omega_A \cong \mathrm{Hom}_k(A, k)$, but the isomorphism is not canonical. The module $\mathrm{Hom}_k(A, k)$ comes equipped with a natural map $\eta_k : \mathrm{Hom}_k(A, k) \to k$, namely evaluation at $1 \in A$. Given an isomorphism $\varphi : \omega_A \to \mathrm{Hom}_k(A, k)$, we get a map $\eta_k\varphi : \omega_A \to k$, and the isomorphism φ may be expressed in terms of this map by the formula $\varphi(w)(a) = (\eta_k\varphi)(aw)$. From this it is easy to see that $\eta_k\varphi$ must be nonzero on the socle of ω_A, since φ would otherwise kill the map that is the inclusion of the socle.

Conversely, given any k-linear map $\eta : \omega_A \to k$, let $\psi_\eta : \omega_A \to \mathrm{Hom}_k(A, k)$ be defined, for $w \in \omega_A$ and $a \in A$, by the formula $\psi_\eta(w)(a) = \eta(aw)$. Show that the map ψ_η is a map of A-modules, and that the following are equivalent:

a. ψ_η is an isomorphism.

b. η generates $\mathrm{Hom}_k(\omega_A, k) \cong A$ as an A-module.

c. η is nonzero on the socle of ω_A.

A map $\eta : \omega_A \to k$ satisfying conditions a–c is called a **residue map**. The name comes from duality theory on a smooth curve over $\mathbf{C}$. There the analogue of the elements of the canonical module are certain meromorphic differential forms, representing classes in the first cohomology group of the cotangent bundle; and the classical residue map is just the sum of the residues of these forms at all points where they have poles. See Kunz [1991, 1992] for some elementary and unusual examples of the use of residue theory on curves.

Having explicit residue maps is often quite useful, and there is an interesting literature on constructing them. The result of Eisenbud and Levine cited above is an application of such results.

Higher Dimension

Exercise 21.11 (Semigroup rings): Here is a case where the computation of ω_A is quite direct. Let $\Gamma \subset \mathbf{N}$ be a "numerical semigroup" —that is a subset containing 0 and closed under addition. Suppose that Γ contains all numbers from a certain number c on, and that c is the smallest number with

this property; c is called the **conductor**. We may use Γ to define a subring of the power series ring in one variable, $A := k[[\{t^\gamma | \gamma \in \Gamma\}]] \subset k[[t]]$. We abbreviate this as $A = k[[t^\Gamma]]$. A good general reference for the properties of such rings, and their relation to one-dimensional complete domains in general, is Herzog and Kunz [1971]. Show that:

a. A is a one-dimensional local domain (and is thus Cohen-Macaulay) with quotient field $k((t))$. The integral closure of A is $A' = k[[t]]$. The ideal $\mathrm{Ann}_A(A'/A)$ (also called the **conductor**) is the ideal $(t^c, t^{c+1}, t^{c+2}, \ldots) \subset A$.

b. Let ω be the vector space spanned by $t^{-\alpha}$ for all integers $\alpha \notin \Gamma$. Show that ω is an A-submodule of $k((t))$, and $\omega \supset tk[[t]]$. For example, if Γ is the semigroup generated by 3, 5, and 6, show that $c = 8$ and that ω is spanned by $t^{-4}, t^{-2}, t^{-1}, t, t^2, t^3, \ldots$.

c. Let $s = t^c$, so that $k[[t]]$ and A are finite modules over $k[[s]]$. By Theorem 21.15, $\omega_A \cong \mathrm{Hom}_{k[[s]]}(A, k[[s]])$. Show that

$$\mathrm{Hom}_{k[[s]]}(A, k[[s]]) = \{\varphi \in \mathrm{Hom}_{k((s))}(k((t)), k((s))) | \varphi(A) \subset k[[t]]\}.$$

d. Show that $\mathrm{Hom}_{k((s))}(k((t)), k((s)))$ is generated by the map σ sending t^i to 0 if i is not a multiple of c and to $s^{i/c}$ if i is a multiple of c. Show that σ generates $\mathrm{Hom}_{k((s))}(k((t)), k((s)))$ as a $k((t))$ module, where as usual the module structure is given by $f\sigma(g) = \sigma(fg)$. Deduce that $\omega_A \cong \{f \in k((t)) | f\sigma(A) \subset k[[s]]\}$. Using this description, show that $\omega \cong \omega_A$.

e. Show that, for any semigroup Γ as above, if for $\gamma \in \Gamma$ then $c - 1 - \gamma \notin \Gamma$.

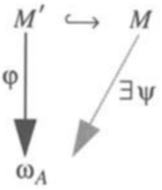
f. Deduce from part d that $\omega_A \cong A$, so that A is Gorenstein, iff Γ satisfies the condition:

$$\gamma \in \Gamma \Leftrightarrow c - 1 - \gamma \notin \Gamma, \text{ or equivalently}$$
$$\mathrm{card}\{\alpha \in [0, 1, \ldots, c-1] | \alpha \notin \Gamma\} = \mathrm{card}\{\alpha \in [0, 1, \ldots, c-1] | \alpha \in \Gamma\}.$$

Semigroups Γ satisfying this condition are called **symmetric** semigroups.

g. Check using these ideas that $k[[t^2, t^5]]$ is Gorenstein, but $k[[t^3, t^4, t^5]]$ is not Gorenstein.

Exercise 21.12: Show that ω_A is an injective in the category of maximal Cohen-Macaulay A-modules in the sense that if $0 \to M' \to M \to M'' \to 0$ is a short exact sequence of maximal Cohen-Macaulay modules, and $\varphi : M' \to \omega_A$ is any homomorphism, there is a homomorphism $\psi : M \to \omega_A$ extending φ:



Exercise 21.13: Show that condition c in Theorem 21.8 may be replaced by the condition

c$'$. $\mathrm{Ext}^d_A(k, \omega_A) \cong k$.

Exercise 21.14 (Cohen-Macaulay type): Suppose that A is a local Cohen-Macaulay ring. The following provides a measure of how far A is from being Gorenstein:

If $x_1, \ldots, x_d$ is a regular sequence in A, the number of generators of the socle of $A/(x_1, \ldots, x_d)$ is called the **(Cohen-Macaulay) type** of A (with respect to $(x_1, \ldots, x_d)$).

a. Show that if A has a canonical module, then the type of A is the minimal number of generators of the canonical module. Deduce that the type does not depend on the regular sequence chosen, and that if P is a prime ideal of A, then type $A_P \leq$ type A.

b. If A is a homomorphic image of a regular local ring R, identify the type of A with an invariant of the resolution of A as an R-module.

c. Show that the type of A is the same as the type of the completion of A—which always has a canonical module.

Exercise 21.15: Here is another construction of Gorenstein rings. It is not as canonical as that of Theorem 21.6 because of the dependence on the regular sequence $(x_1, \ldots, x_c)$, but it is often simpler computationally:

Let R be a regular local ring and let $x_1, \ldots, x_c$ be a regular sequence in R. Let $y \in R, y \notin (x_1, \ldots, x_c)$ and set $J := ((x_1, \ldots, x_c) : y)$. Prove that J has codimension c and R/J is Gorenstein. If $J \neq (x_1, \ldots, x_c)$, show that y satisfies the formula $(y, x_1, \ldots, x_c) = ((x_1, \ldots, x_c) : J)$, so J determines y modulo $(x_1, \ldots, x_c)$ up to units.

Exercise 21.16:* Let $A = k[x_1, \ldots, x_r]/(f_1, \ldots, f_c)$, where k is a field, $k[x_1, \ldots, x_r]$ is a polynomial ring where the variable x_i has degree ε_i, and f_j is a form of degree δ_j such that $f_1, \ldots, f_c$ form a regular sequence. Show directly from the definition of a graded canonical module that

$$\omega_A = A(-\Sigma\varepsilon_i + \Sigma\delta_j).$$

Exercise 21.17 (Hilbert series of a graded Cohen-Macaulay ring): Suppose that k is a field and that $A = k \oplus A_1 \oplus \cdots$ is a positively

graded Cohen-Macaulay ring. Let $h_A(t) = 1+(\dim_k A_1)t+(\dim_k A_2)t^2+\cdots$ be the Hilbert series of A. Similarly, we write $h_{\omega_A}(t)$ for the Hilbert series of the graded module ω_A.

a. If $\dim A = 0$, then $h_A(t)$ is a polynomial; suppose it has degree n. Show that $h_{\omega_A}(t) = t^n h_A(t^{-1})$.

b. Suppose that A is Cohen-Macaulay of dimension d, and that $f_1, \ldots, f_d$ is a regular sequence of homogeneous elements in A with degrees $\delta_1, \ldots, \delta_d$. Set $B = A/(f_1, \ldots, f_d)$. Show that $h_A(t) = h_B(t)\Pi_{i=1}^d(1 - t^{\delta_i})$.

c. In the situation of part b, show that for some integer n we have $h_{\omega_A}(t) = (-1)^{\dim A} t^n h_A(t^{-1})$. In particular, if A is Gorenstein, $h_A(t)/h_A(t^{-1}) = (-1)^{\dim A} t^n$ for some n.

d. Show that the converse of the assertion in part c is false, even in the zero-dimensional case, by computing the Hilbert series of the ring

$$k[x, y]/(x^2, xy, y^3)$$

and showing that this ring is not Gorenstein. Remarkably, the converse is true for domains; see Exercise 21.19.

The Canonical Module as Ideal

Exercise 21.18 (When is the canonical module an ideal?):* Let A be a local Cohen-Macaulay ring, and suppose that A has a canonical module ω_A. We have already mentioned that condition c of Theorem 21.8 has the meaning that ω_A is of rank 1. Here we show that if A is a domain, then ω_A is isomorphic to an ideal. In fact, it is not hard to give the exact conditions under which ω_A is isomorphic to an ideal.

a. Prove that ω_A is isomorphic to an ideal of A iff A is generically Gorenstein, in the sense that A_P is Gorenstein for every minimal prime ideal P of A. (Note that since a field is Gorenstein, any domain—even any reduced ring —is generically Gorenstein.)

b. If $A = k \oplus A_1 \oplus \cdots$ is a positively graded Cohen-Macaulay ring, with k a field, and A is generically Gorenstein, prove that ω_A is isomorphic, up to a shift in degree, to a homogeneous ideal of A.

Exercise 21.19 (Stanley's Gorenstein Criterion):* (Stanley [1978]

a. Suppose that $A = k \oplus A_1 \oplus \cdots$ is a positively graded Cohen-Macaulay domain. If A satisfies $h_A(t)/h_A(t^{-1}) = (-1)^{\dim A} t^n$ for some n, then A is Gorenstein.

b. It is not enough to assume that A is a graded generically Gorenstein algebra (that is, locally Gorenstein at the minimal primes) or even that A is reduced. Show that $A := k[x,y,z,w]/(xw, yw, xys)$ is a reduced Cohen-Macaulay ring of dimension 2 with Hilbert series $h_A(t) = (1+2t+t^2)/(1-t)^2$, but A is not Gorenstein.

Exercise 21.20:* Let A be a local Cohen-Macaulay ring. Show that if $I \subset A$ is a proper ideal, and I is a canonical module for A, then I has codimension 1 and A/I is Gorenstein.

Exercise 21.21:* Let A be a local Cohen-Macaulay domain with a canonical module. If A is factorial, prove that A is Gorenstein.

Linkage and the Cayley-Bacharach Theorem

Exercise 21.22 (The graded case of linkage): Suppose that k is a field and that $A = k \oplus A_1 \oplus \cdots$ is a positively graded Gorenstein ring with canonical module $A(\delta)$. Suppose that I and J are ideals of codimension 0 in A that are directly linked in A, and that $B := A/I$ and $C := A/J$ are Cohen-Macaulay. Show that $\omega_C = (0 : I)(\delta)$.

Exercise 21.23 (Resolutions and linkage): There is much more to be said about linkage. For example, when X and Y are curves in $\mathbf{P}^3$, directly linked by equations of degrees d and e, Peskine and Szpiro [1974] prove that the degrees and genera of X and Y are related by the formulas:

$$\begin{aligned}&\deg X + \deg Y = de\\ &\operatorname{genus} X - \operatorname{genus} Y = (\operatorname{degree} X - \operatorname{degree} Y)(d+e-4)/2.\end{aligned}$$

This and many other of the facts about linkage are proved by using resolutions. The technique rests on the following observations:

Suppose that R is a regular local ring, $I \subset R$ is an ideal of codimension c such that R/I is Cohen-Macaulay, and $x_1, \ldots, x_c$ is a regular sequence contained in I. Let $J = ((x_1, \ldots, x_c) : I)$ be the linked ideal. The linkage process can be used to go from a resolution of R/I to a resolution of R/J. Let

$$\mathcal{F} : 0 \to F_c \to \cdots \to F_1 \to R \to R/I \to 0$$

be a free resolution of R/I, and let

$$K : 0 \to \wedge^c R^c \to \cdots \to R^c \to R \to R/(x_1, \ldots, x_c) \to 0$$

be the Koszul complex that resolves $R/(x_1, \ldots, x_c)$. Let $\varphi : K \to \mathcal{F}$ be a map of complexes extending the projection map $R/(x_1, \ldots, x_c) \to R/I$, and let $\mathcal{E}$ be the mapping cone of φ, so that $\mathcal{E}$ has the form

$$\mathcal{E} : 0 \to \wedge^c R^c \to \wedge^{c-1} R^c \oplus F_c \to \cdots \to R^c \oplus F_2 \to R \oplus F_1 \to R.$$

a. Identifying $\wedge^i R^{c*}$ with $\wedge^{c-i} R^c$, show that the dual

$$\mathcal{E}^* : 0 \to R \to (R \oplus F_1)^* \to (R^c \oplus F_2)^* \to \cdots \to R^c \oplus F_c^* \to R$$

is a free resolution of R/J by proving that

$$R/J = \mathrm{Ext}_R^c(I/(x_1, \ldots, x_c), R)$$

and deducing an exact sequence

$$\begin{aligned} &0 \to \mathrm{Ext}_R^c(R/I, R) \to \mathrm{Ext}_R^c(R/(x_1, \ldots, x_c), R) \\ &\mathrm{Ext}_R^c(I/(x_1, \ldots, x_c), R) \to 0 \end{aligned}$$

from the exact sequence

$$0 \to I/(x_1, \ldots, x_c) \to R/(x_1, \ldots, x_c) \to R/I \to 0.$$

b. The resolution in part a is never minimal —it does not even have the right length. Show that it contains a resolution of the right length (but in general is still not minimal) of the form

$$0 \to (F_1)^* \to (R^c \oplus F_2)^* \to \cdots \to R^c \oplus F_c^* \to R.$$

If the elements $x_1, \ldots, x_c$ form part of a minimal set of generators for I, then the map $\varphi_1 : R^c \to F_1$ is a split inclusion, and there is a resolution of R/J of the form

$$0 \to (F_1/\varphi_1(R^c))^* \to (F_2)^* \to \cdots \to R^c \oplus F_c^* \to R.$$

c. Deduce from this another proof of the statement that if I is principal modulo $x_1, \ldots, x_c$, then R/J is Gorenstein.

d. Now suppose that R/I is Gorenstein, and identify F_c with R. The map $\varphi : K \to \mathcal{F}$ has as its degree-c part a map $\varphi_c : R \to F_c = R$; that is, an element $s \in R$. Show that s generates the ideal $(I : (x_1, \ldots, x_c))$.

e. In the special case where $I = (f_1, \ldots, f_c)$, generated by a regular sequence, let (a_{ij}) be a $c \times c$ matrix such that $\Sigma a_{ij} f_i = x_j$. We may choose φ so that φ_0 is the identity and φ_1 is given by the matrix (a_{ij}). If $R/(x_1, \ldots, x_c)$ is zero-dimensional, and I is the maximal ideal of R, show that $\det(a_{ij})$ generates the socle of $R/(x_1, \ldots, x_r)$.

Exercise 21.24 (Cayley-Bacharach): The classical statement of the Cayley-Bacharach theory, due to Chasles (who was generalizing Pascal's theorem in projective geometry) is the following: Any cubic vanishing on 8 of the 9 points in which two cubic curves meet in the plane vanishes on the ninth point. Cayley and Bacharach generalized this to plane curves of higher degree. The most general version is perhaps the statement that a complete intersection is Gorenstein. Connect these rather dissimilar statements as follows:

a. Intrinsic version: Suppose that A is a graded one-dimensional Gorenstein ring, with $A_0 = k$ a field, and $x \in A_1$ is a nonzerodivisor. For any graded A-module N, finite-dimensional over k in each degree, write

$$\mathrm{Hom}_{\mathrm{gr}}(N, k) = \oplus_n \mathrm{Hom}_k(N_n, k)$$

for the "graded dual" of N. Let M be the cokernel of the localization map $A \to A[x^{-1}]$. Show that M is a graded module, finite-dimensional in each degree. Set $\omega = \mathrm{Hom}_{\mathrm{gr}}(M, k)$. Show that $\omega \cong \omega_A$, the canonical module, as follows:

Dualize the exact sequence

$$0 \to A \to A[x^{-1}] \to M \to 0$$

to get an exact sequence

$$0 \to \omega \to \mathrm{Hom}_{\mathrm{gr}}(A[x^{-1}], k) \to \mathrm{Hom}_{\mathrm{gr}}(A, k) \to 0.$$

Deduce from this exact sequence that x is a nonzerodivisor on ω and that

$$\omega/x\omega \cong \{\varphi \in \mathrm{Hom}_{\mathrm{gr}}(A, k) | x\varphi = 0\}.$$

Show that this last is isomorphic to $\mathrm{Hom}_k(A/xA, k)$, the canonical module of A/xA.

Show that if A is the homogeneous coordinate ring of a set Γ of d points in projective space, then $\dim_k M_n$ is the failure of Γ to impose independent conditions on forms of degree n —that is, the difference between the number of points in Γ and the codimension of the space of forms of degree n vanishing on Γ in the space of all forms of degree n.

b. Numerical version: Let Γ be a set of d points (or a subscheme, for those who know what that is) in $\mathbf{P}^r$ whose ideal is a complete intersection $(f_1, \ldots, f_r)$ of forms, with $\deg f_i = d_i$. Let Γ' be a subset of Γ, and let Γ'' be the residual set. Show that $I_{\Gamma''} = (I_\Gamma : I_{\Gamma'})$, and conversely, as in Theorem 21.23. (In the scheme case, define Γ'' by this equation.) Apply part a to the ring $k[x_1, \ldots, x_r]/(f_1, \ldots, f_r)$ to show that the dimension of the space of forms of degree m vanishing on Γ', modulo those vanishing on Γ, is equal to the failure of Γ'' to impose independent conditions on forms of degree $e - m$, where $e = \Sigma d_i - r - 1$.

c. Basic application: Taking Γ'' to be a single point, show that Γ'' imposes independent conditions on forms of degree 0. Deduce that any form of degree $\Sigma d_i - r$ vanishing on all but one of the points of Γ vanishes on all of Γ. The preceding statement about cubics is a special case.

Appendix 1
Field Theory

In this appendix we prove some results about infinite field extensions that are used in the text. We assume that the reader is familiar with algebraic field extensions, and with the notions of separability and inseparability for such extensions. The necessary background (and more) can be found in Lang [1993] or nearly any other basic algebra text. A general source for the material in this appendix is Bourbaki [1981] Chapter 5.

A1.1 Transcendence Degree

Let X be an affine variety over a field k. The most primitive notion of the dimension of X is the number of independent rational functions on X, that is, the transcendence degree over k of the field $K(X)$ of rational functions on X. Here we will show algebraically that this is well defined.

The simplest kind of field extension is a pure transcendental extension defined as follows:

Definitions. *Let $\{x_b\}_{b\in B}$ is a set of indeterminates. The field $k(\{x_b\}_{b\in B})$ of rational functions in the indeterminates x_b is called a* ***pure transcendental extension*** *of k.*

If $k \subset K$ are fields, and $B \subset K$ is a set of elements, then B is ***algebraically independent over k*** *if there is a homomorphism $k(\{x_b\}_{b\in B}) \to K$ sending x_b to b.*

Homomorphisms of fields are automatically monomorphisms. Thus if B is algebraically independent over k the subfield $k(B)$, generated by B, is

isomorphic to $k(\{x_b\}_{b\in B})$. The set B is a ***transcendence basis*** for K over k if B is algebraically independent and K is algebra over the subfield $k(B)$.

The reader may verify that the following more elementary description of algebraic independence is equivalent to the one just given: The set B is algebraically independent over k if for any integer n, any nonzero polynomial $f(t_1, \dots, t_n)$ with coefficients in k, and any set $b_1, \dots, b_n$ of distinct elements of B we have $f(b_1, \dots, b_n) \neq 0$.

Theorem A1.1. *Let $k \subset K$ be fields. If B and B' are transcendence bases of K over k, then B and B' have the same cardinality.*

We shall give the proof only in the case where B and B' are finite. The proof uses the **exchange property**: Suppose we are given a set L (in the proof L will be a finite set such that K is algebraic over $k(L)$) and a nonempty distinguished class of finite subsets $\mathcal{B}$ called **bases**, no one contained in another. We say that $\mathcal{B}$ has the **exchange property** if given $B, B' \in \mathcal{B}$ and $b' \in B'$, there is an element $b \in B$ such that the set $B' \cup \{b\} - \{b'\}$ is again a basis. A set L with a family of bases $\mathcal{B}$ satisfying the exchange property is called a **matroid**. (See White [1986], [1987], and [1992] for a sense of myriad occurrences of this notion.) The collection of bases in a finite dimensional vector space is a matroid, and we shall see that the set of transcendence bases of a given field extension of finite transcendence degree is a matroid too. (The same is true for the p-bases that we will introduce below, but we will not need this.)

The exchange property ensures that all the sets in $\mathcal{B}$ have the same number of elements:

Lemma A1.2. *If $\mathcal{B}$ is a matroid, then all the bases in $\mathcal{B}$ have the same cardinality.*

Proof. Let r be the minimal cardinality of a basis. If B is a basis, we prove that B has r elements by downward induction on the number of elements that B shares with some basis B' of cardinality r. To start the induction, observe that if $\operatorname{card}(B \cap B') = r$, then $B = B'$ since no basis can properly contain another. If $\operatorname{card}(B \cap B') = s < r$, then there is an element $b' \in B'$ that is not in B. By the exchange property there is a basis of the form $B' - \{b'\} \cup \{b\}$ for some $b \in B$. This basis has cardinality r and shares $s+1$ elements with B, so we are done by induction. □

Proof of Theorem A1.1 in the case of finite transcendence. We shall show that the set $\mathcal{B}$ of finite subsets of L that are transcendence bases for K over k has the exchange property. By Lemma A1.2 this implies the conclusion of the theorem.

Let $B = \{b_1, \dots, b_s\}$ and $B' = \{b'_1, \dots, b'_r\}$ be two transcendence bases of K over k. Since K is algebraic over $k(b'_1, \dots, b'_r)$, there is for each i an irreducible polynomial $p_i \in k(b'_1, \dots, b'_r)[Y]$ such that $p_i(b_i) = 0$. Multiplying

by a suitable polynomial in $b'_1, \ldots, b'_r$, we may assume that the coefficients of each p_i are polynomials in the b'_i. The element $b'_1 \in K$ is algebraic over $k(b_1, \ldots, b_s)$. If none of the p_i involves b'_1 nontrivially, then $b_1, \ldots, b_s$ are algebraic over $\{b'_2, \ldots, b'_r\}$, and thus b'_1 is algebraic over $\{b'_2, \ldots, b'_r\}$, contradicting our hypothesis that B'_1 is a transcendence basis. Thus some p_i involves b'_1.

For this value of i, we claim that $\{b_i, b'_2, \ldots, b'_r\}$ is a transcendence basis. The polynomial relation $p_i(b_i) = 0$ shows that b'_1 is algebraic over $\{b_i, b'_2, \ldots, b'_r\}$. If $\{b_i, b'_2, \ldots, b'_r\}$ were algebraically dependent, then since $\{b'_2, \ldots, b'_r\}$ is algebraically independent, b_i would be algebraically dependent on $\{b'_2, \ldots, b'_r\}$. Thus b'_1 would be algebraically dependent on $\{b'_2, \ldots, b'_r\}$, contradicting the hypothesis that $\{b'_1, b'_2, \ldots, b'_r\}$ is a transcendence basis. □

A1.2 Separability

Let $k \subset K$ be fields. If K is algebraic over k, we say that K is separable over k if for every $\alpha \in K$ the minimal polynomial satisfied by α over k has no multiple roots in an algebraic closure of k. In characteristic 0 every algebraic extension is separable. We shall explain the natural extension of separability to nonalgebraic extensions, something that is useful in the theory of coefficient fields of complete rings and in the study of the module of differentials. Nearly all of what follows is from MacLane [1939]. (This very readable paper contains much more information as well.)

Definition. *K is **separably generated** over k if there exists a transcendence base $\{x_\lambda\}_{\lambda\in\Lambda}$ for K such that K is a separable algebraic extension of $k(\{x_\lambda\}_{\lambda\in\Lambda})$. K is **separable** over k if every subfield of K that is finitely generated over k is separably generated over k.*

It is not quite obvious, but we shall soon see that a separably generated extension field is separable.

In characteristic 0 the fact that every algebraic extension is separable makes it clear that every extension is separably generated, and thus that every extension is separable. Thus we will assume for the remainder of this section that k is a field of characteristic $p \neq 0$.

We write k^{1/p^∞} for the union, over all n, of the field generated by the (p^n)th roots of elements of k. It turns out that the separability of K over k depends on the relationship between K and k^{1/p^∞}.

If K and L are subfields of a field K', then we write $L * K$ for the **compositum of L and K**, that is, the field generated by L and K. We say that L and K are **linearly disjoint** over a common subfield k if the multiplication map $L \otimes_k K \to L * K$ taking $a \otimes b$ to ab is an isomorphism. (The name comes from the remark that L and K are linearly disjoint iff

for every set $\{x_i\}$ of elements of L linearly independent over k, the set $\{x_i\}$ remains linearly independent over K.)

Theorem A1.3 (MacLane [1939]). *The following statements are equivalent:*

a. K is separable over k.

b. For every field extension L of k, the ring $L \otimes_k K$ is reduced.

c. $k^{1/p^\infty} \otimes_k K$ is reduced.

d. k^{1/p^∞} is linearly disjoint from K.

Proof. a $\Rightarrow$ b: Since L is flat over k, we may write $L \otimes_k K = \cup_{K'} L \otimes_k K'$, where K' runs over all subfields of K finitely generated over k. Thus it is enough to do the case where K is finitely generated over k, and we may assume that K is separably generated over k. To show that K satisfies condition b, it is enough to show that both purely transcendental extensions and separable algebra extensions satisfy condition b.

The purely transcendental case is easy since $L \otimes_k k(x_1, \ldots, x_r) = L(x_1, \ldots, x_r)$, the field of rational functions over L. For the algebra case, it suffices, by the same argument about unions as before, to treat the case where $K = k(\alpha)$ is generated by one algebra element α.

Suppose α satisfies a minimal polynomial $f(x)$. We have $K = k[x]/(f(x))$, so $L \otimes_k K = L[x]/(f(x))$. As f splits into a product of distinct linear factors over the algebra closure of L, it splits into a product of relatively prime factors over L, say $f = \Pi_i f_i$. By the Chinese remainder theorem (Proposition 2.13 or Exercise 2.6) we have $L[x]/(f(x)) = \Pi_i L[x]/(f_i(x))$, a product of fields. This shows that $L \otimes_k K$ is reduced.

b $\Rightarrow$ c: Obvious.

c $\Rightarrow$ d: Because K is flat over k, we have $L \otimes K = U_L\ L \otimes K$, where L runs over the subfields $L \subset k^{1/p^\infty}$ finite over k, so it suffices to show that K is linearly disjoint from such an L.

We do induction on the degree of L over k. As L may be obtained by successively adjoining pth roots, we may choose a subfield $L_1 \subset L$, containing k, such that $L = L_1(\alpha)$ for some α with $\alpha^p \in L_1$ and $\alpha \notin L_1$. As $L \otimes_k K = L \otimes_{L_1} (L_1 \otimes_k K)$, and as $k^{1/p^\infty} \otimes_k K = k^{1/p^\infty} \otimes_{L_1} (L_1 \otimes_k K)$, we may assume that $L_1 = k$.

Since $\alpha^p \in k$, the order of the extension $K(\alpha)$ is either 1 or p, so it suffices to show that $\alpha \notin K$. But if $\alpha \in K$, then $x^p - \alpha$, the irreducible equation of α over k, would split as $(x - \alpha^{1/p})^p$ in K, and thus $L \otimes_k K = K[x]/(x^p - \alpha) = K[x]/(x - \alpha^{1/p})^p$ is not reduced. Since $L \otimes_k K \subset k^{1/p^\infty} \otimes_k K$, this contradicts our hypothesis and shows that L and K are linearly disjoint.

To prove that d $\Rightarrow$ a, we need a method for recognizing separating transcendence bases. We shall use a more general notion, which also plays an important role in the study of coefficient fields of complete local rings.

A1.3 p-Bases

Definition (Teichmüller [1936]). *If $k \subset K$ are fields of characteristic p, then a collection of elements $\{x_\lambda\}_{\lambda\in\Lambda} \subset K$ is a* **p-basis** *for K over k if the set W of monomials in the x_λ having degree $< p$ in each x_λ separately forms a vector space basis for K over the subfield $k * K^p$.*

We next show that every extension K of k has a p-basis. It will be useful to know that if K' is any K^p-algebra contained in K, then K' is a field. This follows from the fact that every element of K' is integral over the field K^p (use Corollary 4.17 or Exercise 4.3). Moreover, if $x \in K$ is not in K', then $K'[x]$ is a field extension of degree p over K'. For the minimal polynomial $f(t)$ satisfied by x over K' must divide $t^p - x^p = (t-x)^p$ and thus has the form $f(t) = (t-x)^n$ for some n. It follows that both x^n and x^p are in K'. If $n < p$, then since n and p are relatively prime, we would have $x \in K'$, contradicting the hypothesis $x \notin K'$.

By Zorn's lemma we may choose a maximal subset $\{x_\lambda\}_{\lambda\in\Lambda} \subset K$ having the property that the elements of the set W defined above are linearly independent over $k * K^p$. The span of W over $k * K^p$ is equal to the subring $k * K^p[B]$ generated by B over $k * K^p$. By the remark above, $k * K^p[B]$ is a subfield of K. We must show that $K = k * K^p[B]$. If this were not the case, then an element x of K not in $k * K^p[B]$ would generate a field extension of degree p over $k * K^p[B]$. Thus $1, x, \ldots, x^{p-1}$ are linearly independent over $k * K^p[B]$. This contradicts the maximality of $\{x_\lambda\}_{\lambda\in\Lambda}$ and shows that $K = k * K^p[B]$. Thus, $\{x_\lambda\}_{\lambda\in\Lambda}$ is a p-basis.

This argument also shows that being a p-basis is equivalent to being a minimal set of generators for K as a field over $k * K^p$. (Since K is algebraic over $k * K^p$, we could also say a minimal set of generators as an algebra over $k * K^p$.)

We need the following facts about p-bases:

Theorem A1.4 (MacLane [1939]). *Suppose that $k \subset K$ are fields of characteristic p. Let B be a p-basis for K over k, and let $q = p^n$ for some n. Let W_q be the set of monomials in the elements of B having degree $< q$ in each element of B.*

a. $K = k * K^q[B]$; that is, W_q spans K as a vector space over $k * K^q[B]$.

b. If K is separable over k, then W_q is a vector space basis for K over $k * K^q[B]$.

c. If K is separable over k, then the elements of B are algebraically independent over $k * K^{p^\infty}$.

Proof. a and b. We do induction on n. The case $n = 1$ follows from the definition of a p-basis, so we may suppose $n > 1$. Let $q' = p^{n-1}$. By induction $K = k * K^{q'}[B]$.

Since raising to the pth power is a ring homomorphism in characteristic p, we may raise this equation to the pth power, getting $K^p = k^p * K^q[B^p]$. Thus $K = k * K^p[B] = k * k^p * K^q[B^p][B] = k * K^q[B]$. This proves part a.

For part b we must show that the elements of W_q are linearly independent. Equivalently, we must show for every finite subset $\{b_1, \ldots, b_s\} \subset B$ that the field $k * K^q[b_1, \ldots, b_s]$ has degree q_s over $k * K^q$. Consider the chain of fields

$$k * K^q \overset{(1)}{\subset} k * K^q[b_1^p, \ldots, b_s^p] \overset{(2)}{\subset} k * K^q[b_1, \ldots, b_s].$$

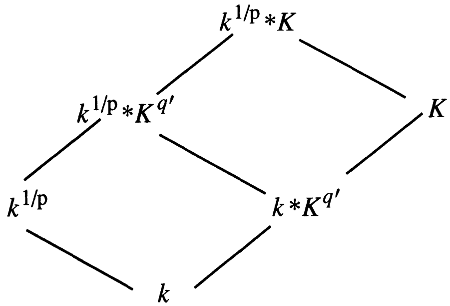
We shall show that the extension (1) has degree $\geq (q')^s$ and that the extension (2) has degree $\geq p^s$. It will follow that the whole extension has degree $\geq q^s$.

Consider first the extension (2). Since $k * K^p[b_1, \ldots, b_s]$ has degree p^s over $k * K^p$, and $k * K^q[b_1^p, \ldots, b_s^p]$ is contained in $k * K^p$, it follows that

$$k * K^q[b_1, \ldots, b_s] = k * K^q[b_1^p, \ldots, b_s^p][b_1, \ldots, b_s]$$

has degree $\geq p^s$ over $k * K^q[b_1^p, \ldots, b_s^p]$.

Next consider the extension (1). By induction $k * K^{q'}[b_1, \ldots, b_s]$ has degree $(q')^s$ over $k * K^{q'}$. Since raising to the pth power is an isomorphism of K onto its image K^p, we see that $k^p * K^q[b_1^p, \ldots, b_s^p]$ has degree $(q')^s$ over $k^p * K^q$. To show that $k * k^p * K^q[b_1^p, \ldots, b_s^p] = k * K^q[b_1^p, \ldots, b_s^p]$ has degree $(q')^s$ over $k * k^p * K^q = k * K^q$, it now suffices to show that $k * K^q$ and $k^p * K^q[b_1^p, \ldots, b_s^p]$ are linearly disjoint over $k^p * K^q$. Since the second field is contained in K^p, it suffices to show that $k * K^q$ and K^p are linearly disjoint over $k^p * K^q$. This is equivalent to the statement that $k^{1/p} * K^{q'}$ and K are linearly disjoint over $k * K^{q'}$, since the pth power map provides an isomorphism between these two situations. Since K is separable over k, the fields $k^{1/p}$ and K are linearly disjoint over k. (These fields are shown in the illustration.) Since



$k * K^{q'}$ is a subfield of K, this implies that $k^{1/p} * K^{q'} = k^{1/p} \otimes_k (k * K^{q'})$, and thus $k^{1/p} * K^{q'}$ and k are linearly disjoint over $k * K^{q'}$ as required. This proves part b.

Part c is an immediate consequence of part b: Since $k * K^{p^\infty} = \bigcap_{q=p^n} k * K^q \subset k * K^q$ for every q, the elements of W_q are linearly independent over $k * K^{p^\infty}$ for every q. That is, no polynomial in the b_i with

coefficients in $k * K^{p^\infty}$ can be zero, so the elements of B are algebraically independent over $k * K^{p^\infty}$ as claimed.

As a consequence we can exhibit the relation between p-bases and separating transcendence bases:

Corollary A1.5. *Let $k \subset K$ be fields of characteristic p.*

a. Any separating transcendence basis for K over k is a p-basis.

b. If B is a p-basis for K over k and K is finitely generated over k, then K is finite and separable over $k(B)$. If in addition K is separable over k, then B is a separating transcendence basis for K over k.

Part b fails for extensions that are not finitely generated, as may be seen from the example at the end of this appendix.

Proof. a. Let B be a separating transcendence basis for K over k, so that K is a separable algebraic extension of $k(B)$. K is a purely inseparable extension of $k(B) * K^p$, so we must have $k(B) * K^p = K$, and we need only show that B is a minimal generating set for K over $k * K^p$. Were this not the case, we could write $B = \{x\} \cup B'$ in such a way that $x \in k(B') * K^p$, so $K = k(B') * K^p$. Since K is separable over $k(B)$, it follows that K^p is separable over $k(B^p)$, and thus $K = k(B') * K^p$ is separable over $k(B')(x^p)$. As the elements of B are algebraically independent over k, we see that $x \notin k(B')(x^p)$, so x is not separable over $k(B')(x^p)$; the contradiction shows that B is p-independent.

b. Suppose that K is finitely generated over k, and let B be a p-basis. Suppose that $x_1, \dots, x_r$ is a transcendence basis of K over $k(B)$, and let K' be the maximal separable extension of $k(B)(x_1, \dots, x_r)$ inside K. Since K is purely inseparable and finitely generated over K', we have $K^{p^n} \subset K'$ for sufficiently large n. But since B is a p-basis, $K = k(B) * K^p$ and thus $K = k(B) * K^{p^n}$ for every n. It follows that $K' = K$.

Since now K is a separable extension of $k(B)(x_1, \dots, x_r)$, we see that K^p is a separable extension of $(k(B)(x_1, \dots, x_r))^p = k(B)^p(x_1^p, \dots, x_r^p)$. It follows that $k(B) * K^p$ is a separable extension of $k(B)(x_1^p, \dots, x_r^p)$. From this we get a contradiction if $r > 0$, since x_1 would not be separable over $k(B)(x_1^p, \dots, x_r^p)$. Thus $r = 0$, so K is algebraic and separable over $k(B)$.

If K is separable over k, then by Theorem A1.4b the elements of B are algebraically independent over $k * K^{p^\infty}$ and thus over k.

Conclusion of the Proof of Theorem A1.3. d $\Rightarrow$ a: Since L is flat over K, we have $L \otimes_k K = \cup_{K'} L \otimes_k K'$, where K' runs over the finitely generated extensions of K. It suffices to prove that such a K' has a separating transcendence basis. Since $K' \subset K$, K' is separable over k. By Corollary A1.5 any p-basis for K' over k is a separating transcendence basis. □

Using Theorem A1.3, we can easily derive some basic facts about separability.

Corollary A1.6. *If K is separable over k, and $k \subset K' \subset K$, then K' is separable over k. In particular, any separably generated field is separable.*

Proof. Use the characterization in part b of Theorem A1.3 and the fact that (with the notation there) $L \otimes K' \subset L \otimes K$. The second statement follows from Corollary A1.5b. $\square$

Corollary A1.7. *If k is a perfect field and K is any field containing k, then K is separable over k.*

Proof. To say that k is perfect means that $k^{1/p^\infty} = k$, so the result is immediate from the characterization of separability in part c of Theorem A1.3. $\square$

Example. A typical example of a field extension that is separable but not separably generated is the following: Let k be any field of characteristic p, let x be an indeterminate, and let $K = \cup_n k(x^{1/p^n})$. Any finitely generated subfield of K containing k is contained in some $k(x^{1/p^n})$, which is purely transcendental over k and thus separable. Consequently K is separable over k. The transcendence degree of K over k is 1, so a transcendence basis for K over k consists of a single element y. But $k(y) \subset k(x,y) \neq k(y, x^{1/p^n})$ for large n; so K contains a purely inseparable extension of $k(y)$, and y is not a separating transcendence basis. If k is perfect then $K^p = K$, so the empty set is a p-basis for K over k, even though K is not finite over k.

A1.3.1 Exercises

Exercise A1.1:* Suppose that R is an integral domain containing a field k, and let S be any reduced ring containing k. As usual, we write $K(R)$ for the quotient field of R. Generalize Theorem A1.3 to show that if $K(R)$ is separable over k then $R \otimes_k S$ is a reduced ring. As an application, prove that if $k \subset K \subset K'$ are fields such that K is separable over k and K' is separable over K, then K' is separable over k.

Exercise A1.2:* Suppose that R and S are integral domains containing a field k. Show that:

a. If $K(R)$ is separable over k and k is algebraically closed in $K(R)$, then $R \otimes_k S$ is a domain. For example, if k is algebraically closed, and R and S are arbitrary domains containing k, then $R \otimes_k S$ is a domain. A special case of this is usually stated in algebraic geometry as: the product of two irreducible varieties over an algebraically closed field is again an irreducible variety.

b. If $K(R)$ is a finite purely inseparable extension of k then $R \otimes_k S$ is irreducible in the sense that it has just one minimal prime.

Exercise A1.3: Recall that if $k \subset K$ is an algebraic extension of fields then there is a unique largest field L with $k \subset L \subset K$ such that L is separable over k. (This amounts to saying that if L and L' are both separable over k, then the compositum $L * L'$ is too.) Show that this result may fail for extensions $k \subset K$ that are not algebraic, even for finitely generated extensions of transcendence degree 1.

Appendix 2
Multilinear Algebra

A2.1 Introduction

In this appendix we shall describe the results on tensor products and the symmetric and exterior algebras that we have already used occasionally in the text, and put them into a somewhat larger context. As an application we explain the construction of a family of complexes generalizing the Koszul complex and the complex used in the Hilbert-Burch theorem in Chapter 20. These complexes arise both in commutative algebra and algebraic geometry.

The symmetric algebra of a free module of rank r over a ring R is simply the graded polynomial ring $S := R[x_1, \ldots, x_r]$. It would be superfluous to explain its importance in commutative algebra: It is the *object* of commutative algebra.

The significance of the exterior algebra is more subtle. It appears in commutative algebra in several ways, one of which is in the study of homomorphisms between free modules (matrices): The exterior algebra is implicitly involved in the construction of the determinant, and more generally in the construction of the lower order minors (= subdeterminants) of a matrix. The exterior algebra also appears, in the guise of the Koszul complex, as the minimal free resolution of the "natural" module $R = R[x_1, \ldots, x_r]/(x_1, \ldots, x_r)$. Consequently $\mathrm{Tor}_{R[x_1,\ldots,x_r]}(R, R)$ is isomorphic to an exterior algebra as an R-module. That this identification is natural may be seen from the fact that the algebra and coalgebra structures of the exterior algebra (to be explained below) coincide with the natural algebra and coalgebra structures on $\mathrm{Tor}_{R[x_1,\ldots,x_r]}(R, R)$, explained in Appendix 3.

Some comments on the material in this appendix may help orient the reader. All of the constructions require extensive use of tensor products. Their definition and basic properties are therefore described in the first section.

The exterior and symmetric algebras are best treated together—it turns out that they are both manifestations of a single skew commutative algebra. It is natural to include another algebra, the tensor algebra, as a tool. The second section of this appendix describes the elementary part of the theory of these algebras: "base change," behavior with respect to direct sums, and right exactness. For a categorical view of these results, see Examples a, b, c in Appendix 5.

The fact that the symmetric and exterior algebras behave well with respect to direct sums has the consequence that they are both *Hopf algebras*—essentially this means that their duals are again algebras, and the two algebra structures are related in a simple way. The graded dual of the exterior algebra of a finitely generated free module F turns out to be an exterior algebra on $F^* = \mathrm{Hom}(F, R)$. The graded dual of the symmetric algebra is, however, a new object, called the *divided power algebra* of F^*. In characteristic 0 this is isomorphic to the symmetric algebra of F^*, but in characteristic $p > 0$ the divided power algebra is not even Noetherian. These matters of duality are dealt with briefly in the section on Coalgebra Structures and Divided Powers.

Divided powers—and indeed Hopf algebras—seem first to have been noticed and systematically exploited by the topologists. The abstract definition comes from Cartan [1954], one of a sequence of talks aimed at computations of the homology groups of the Eilenberg-Maclane spaces $K(\pi, n)$. But divided power algebras appear naturally in commutative algebra as well in the construction of free resolutions and elsewhere. (One leading case, that of the minimal free resolution of the residue class field of a local ring, bears a tight analogy with topology; see for example Avramov and Halperin [1986].)

A more comprehensive definition of multilinear algebra than the one we have adopted might include the representation theory of the general linear group GL. For an appendix, however, such a definition would be catastrophic: Even to include the most classical part of the theory, the theory of representations over a field of characteristic 0, alias the theory of symmetric functions, would take at least one whole book! (An excellent recent book including such material is Fulton and Harris [1991].) Moreover, the applications to "standard" commutative algebra actually involve parts of representation theory that are nonclassical—things like integral representation of $\mathrm{GL}(n, \mathbf{Z})$—so the theory of symmetric functions would not be enough. And the state of affairs in this "integral" representation theory is not so good: Many simple things that would be desirable for a natural treatment of the subject are topics of current research, and thus in no state for an introductory text. Since the connection with representation theory

is too significant to ignore completely, we shall sketch the beginnings in the section on Schur Functors.

In the last section we explain some constructions of complexes by multilinear algebra—the Eagon-Northcott and Buchsbaum-Rim complexes and a family of complexes to which they belong.

For a more detailed treatment of the basic material included here, see Bourbaki [1970], Chapter III.

Notation: Throughout this section, R will denote a commutative ring, and M will denote an R-module. Tensor products will be taken over R unless otherwise noted, and we sometimes write $\otimes$ for $\otimes_R$.

The central case of interest is the one where M is a finitely generated free R-module, and though applications of other cases do occur occasionally, we suggest that the inexperienced reader think of this case throughout.

A2.2 Tensor Products

Definition. *Let M and N be R-modules. The* ***tensor product*** *of M and N over R, written $M \otimes_R N$, is the R-module generated by symbols $m \otimes n$ for $m \in M$ and $n \in N$, with relations*

$$\begin{aligned} rm \otimes n &= m \otimes rn \\ (m + m') \otimes n &= m \otimes n + m' \otimes n \\ m \otimes (n + n') &= m \otimes n + m \otimes n'. \end{aligned}$$

These relations say precisely that the natural map $b : M \times N \to M \otimes_R N$ taking (m, n) to $m \otimes n$ is **bilinear**. It follows that the tensor product, together with the map b, has (and is characterized by) the following ***universal property***: For any module P the bilinear maps $M \times N \to P$ are in one-to-one correspondence with the homomorphisms $M \otimes_R N \to P$ by means of composition with b; that is,

$$\begin{aligned} \mathrm{Hom}_R(M \otimes_R N, P) &\cong \{\text{Bilinear maps } M \times N \text{ to } P\} \\ \varphi &\longmapsto \varphi b. \end{aligned}$$

It is elementary that a bilinear map $M \times N \to P$ is the same as a homomorphism $M \to \mathrm{Hom}_R(N, P)$, so we may rewrite the natural isomorphism above as a natural isomorphism

$$\mathrm{Hom}_R(M \otimes_R N, P) \cong \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)).$$

In categorical language this says that the functor $- \otimes_R N$ is the left adjoint of the functor $\mathrm{Hom}_R(N, -)$. See Appendix A5 for the general definition.

Either from general category theory or from simple manipulation of the definition the following facts are easy to prove.

Proposition A2.1 (Properties of $\otimes$). *If M, N, P and Q are R-modules, then*

a. **(Associativity and commutativity)***: $M \otimes_R N \cong N \otimes_R M$ and $M \otimes_R (N \otimes_R P) \cong (M \otimes_R N) \otimes_R P$.*

b. **(Direct sums)***: $(M \oplus N) \otimes_R P \cong (M \otimes_R P) \oplus (N \otimes_R P)$.*

c. **(Right exactness)***: If $M \to N \to P \to 0$ is an exact sequence, then the sequence*

$$Q \otimes_R M \to Q \otimes_R N \to Q \otimes_R P \to 0$$

obtained by tensoring with Q is also exact.

d. **(Base change)***: If R' is an R-algebra and U is an R'-module, then $\mathrm{Hom}_{R'}(R' \otimes_R M, U) \cong \mathrm{Hom}_R(M, U)$ and*

$$U \otimes_R M \cong U \otimes_{R'} (R' \otimes_R M).$$

Proof. Properties a, b, and c are all best checked by using the universal property. For example, the associativity formula in a comes from the fact that a bilinear map (to anywhere) from either $M \times (N \otimes_R P)$ or from $(M \otimes_R N) \times P$ to Q is the same as a trilinear map from $M \times N \times P$ to Q. For statement b note that a bilinear map from $(M \oplus N) \times P$ is the sum of a bilinear map from $M \times P$ and one from $N \times P$. For property c, it is enough to show that the cokernel of the map $Q \otimes_R M \to Q \otimes_R N$ has the same universal property as $Q \otimes_R P$. But maps from the cokernel correspond to bilinear maps from $Q \times N$ that kill the elements of $Q \times M$, and these are the same as bilinear maps from $Q \times P$.

Part d has a slightly different flavor. For the isomorphism $\mathrm{Hom}_{R'}(R' \otimes_R M, U) \cong \mathrm{Hom}_R(M, U)$ we note that the left side may be identified with the set of R-bilinear maps φ of $R' \times M$ to U that satisfy $\varphi(r' \times m) = r'\varphi(1 \times m)$ for any $r' \in R'$. But such a φ is determined by the R-linear map $\psi(m) = \varphi(1 \times m)$, and the same formulas may be read backward, starting from an arbitrary R-linear map ψ.

To prove that $U \otimes_R M \cong U \otimes_{R'} (R' \otimes_R M)$, we give maps in both directions. The map $(u, m) \mapsto u \otimes_{R'} 1 \otimes m$ is R-bilinear, and thus defines a map from left to right. The map sending $(u, \sum s_i \otimes m_i) \in U \times (R' \otimes_R M)$ to $\sum (s_i u \otimes m_i) \in U \otimes_R M$ is R'-bilinear, and defines the inverse map. $\square$

A2.3 Symmetric and Exterior Algebras

We begin by introducing the main characters.

Definitions. *The* ***tensor algebra*** *of the R-module M is the graded, noncommutative algebra*

$$T_R(M) := R \oplus M \oplus (M \otimes_R M) \oplus \cdots,$$

where the product of $x_1 \otimes \cdots \otimes x_m$ and $y_1 \otimes \cdots \otimes y_n$ is $x_1 \otimes \cdots \otimes x_m \otimes y_1 \otimes \cdots \otimes y_n$. In the most interesting case, where M is a free R-module on the x_i, this is simply the free (noncommutative) algebra on the x_i. When the ground ring R is clear from context we shall simply write $T(M)$ for $T_R(M)$.

The **symmetric algebra** of M is the algebra $\boldsymbol{S_R(M)}$ obtained from $T_R(M)$ by imposing the commutative law, that is, by factoring out the two-sided ideal generated by the relations $x \otimes y - y \otimes x$ for all x, $y \in M$. When the context is clear we shall drop the subscript R and write $S(M)$ for the symmetric algebra.

Finally, the **exterior algebra** of M is the algebra $\boldsymbol{\wedge_R(M)}$ obtained from $T_R(M)$ by imposing skew-commutativity, that is, by factoring out the two-sided ideal generated by the elements $x^2 = x \otimes x$ for all $x \in M$. (From the formula $(x+y) \otimes (x+y) = x \otimes x + x \otimes y + y \otimes x + y \otimes y$ we see that $x \otimes y + y \otimes x$ goes to 0 in $\wedge_R M$ for all $x, y \in M$, so that $\wedge_R M$ really is skew-commutative.) When the context is clear we shall drop the subscript R and write $\wedge(M)$ for the exterior algebra.

So far we have ignored the possibility that the module M is graded. We shall now use this possibility to treat the exterior and symmetric algebras together. For the rest of this and the next section we adopt:

Sign Convention and Definition. *R will denote a graded ring concentrated in degree 0 (that is, $R = R_0$), and M will denote a $\mathbf{Z}$-graded R-module, $M = \oplus_{i \in \mathbf{Z}} M_i$. The tensor algebra $T_R(M)$ becomes a $\mathbf{Z}$-graded module with ith graded component*

$$T_R(M)_i = \oplus_{j_1 + \cdots + j_n = i} M_{j_1} \otimes \cdots \otimes M_{j_n}.$$

With this convention in hand, we define the **symmetric algebra**, written $\boldsymbol{\mathcal{S}_R(M)}$ or $\mathcal{S}(\boldsymbol{M})$, to be the graded R-algebra $T_R(M)/I$, where I is the two-sided ideal generated by the "skew-commutativity" relations

$$\text{(SC)} \qquad ab - (-1)^{(\deg a)(\deg b)} ba, \quad a^2 = 0 \text{ if } \deg a \text{ is odd},$$

for homogeneous elements a, b of M.

If M is concentrated in even degrees (that is, $M_i = 0$ for all odd i), then $\mathcal{S}(M) = S(M)$, the symmetric algebra just defined, and is commutative.

If, on the contrary, M is concentrated in odd degrees (that is, $M_i = 0$ for all even i), then $\mathcal{S}(M) = \wedge(M)$ is the exterior algebra on M. When we write $S(M)$ or $\wedge(M)$ here, we shall think of M as an ungraded module, and identify these algebras with $\mathcal{S}(M)$ where M is given degree 2 or 1, respectively.

In general, all homogeneous elements a, b of $\mathcal{S}(M)$ satisfy the skew-commutativity relation, not just those coming from M; this follows at once since every element of $\mathcal{S}(M)$ is a sum of products of elements of M. The elements of even degree in $\mathcal{S}(M)$ commute with every element, while the elements of odd degree anticommute with each other (that is $ab = -ba$). We could have defined $\mathcal{S}$ in terms of S and $\wedge$; see Exercise A2.1.

We define the dth symmetric power of M, written $\mathcal{S}_{R,d}(M)$ or $\mathcal{S}_d(M)$, to be the image in $\mathcal{S}(M)$ of $M \otimes \cdots \otimes M$ (d factors) in $T(M)$. If M is concentrated in even degree, then this is also called $S_d(M)$; if M is concentrated in odd degree, then it is also called the dth exterior power, $\wedge^d(M)$. We have $\mathcal{S}(M) = \oplus_{d \geq 0} \mathcal{S}_d(M)$. Note that each $\mathcal{S}_d(M)$ is a graded module.

We have defined $\mathcal{S}(M)$ as a quotient of the tensor algebra. The reader should be warned that in the classical literature (that is, when R is a field of characteristic 0) one often sees the symmetric and exterior powers identified with the subspaces of symmetric or skew-symmetric tensors (spaces of elements of $T_d(M)$ invariant under appropriate actions of the symmetric group on d letters). It turns out that this identification is possible in the case of the exterior powers of a free module, but it may run into trouble more generally and in particular is wrong for the symmetric powers if R does not contain the field of rational numbers. (When M is a free module, one gets the homogeneous components of the divided power algebra $\mathcal{D}(M)$ defined below.) In any case, these subspaces do not form a subalgebra (instead they form a sub-coalgebra), so the identification cannot be used to define the algebraic structure.

It is perhaps not obvious that the sign convention we have introduced is more than an artifice to save space by talking of two things at once. Actually, it appeared first in topology (the cup product on cohomology is skew commutative in the sense defined here); now it even appears in physics.

Here are some elementary properties of $T(M)$ and $\mathcal{S}(M)$ that are the basis for many computations. Specializing to the case of a module concentrated in even or odd degree, we get corresponding properties for $S(M)$ and $\wedge(M)$; since these are the forms that are commonly used, we have included the statements explicitly:

Proposition A2.2.

a. **(Functoriality)** $\mathcal{S}_1(M) = M$*; and given any map of modules* $\varphi : M \to N$*, there is a unique map of* R*-algebras* $\mathcal{S}(\varphi) : \mathcal{S}(M) \to \mathcal{S}(N)$ *carrying* $\mathcal{S}_1(M) = M$ *to* $\mathcal{S}_1(N) = N$ *via* φ.

b. **(Base change)** *If* R' *is an* R*-algebra and* M *is an* R*-module, then*

$$T_{R'}(R' \otimes_R M) = R' \otimes_R T_R(M),$$
$$\mathcal{S}_{R'}(R' \otimes_R M) = R' \otimes_R \mathcal{S}_R(M),$$

and thus

$$S_{R'}(R' \otimes_R M) = R' \otimes_R S_R(M),$$
$$\wedge_{R'}(R' \otimes_R M) = R' \otimes_R \wedge_R(M).$$

c. **(Direct sums)** $T(M \oplus N) = T(M) \otimes T(N) \otimes T(M) \otimes T(N) \cdots$, *the set of finite sums of tensor products of elements all but finitely many of which are* $1 \in R = T_0(M) = T_0(N)$. *Moreover,*

$$\mathcal{S}(M \oplus N) = \mathcal{S}(M) \otimes \mathcal{S}(N),$$

and thus

$$S(M \oplus N) = S(M) \otimes S(N),$$
$$\wedge(M \oplus N) = \wedge(M) \otimes \wedge(N).$$

d. **(Short exact sequences)** *An exact sequence*

$$K \to N \to M \to 0$$

of R-modules gives rise to exact sequences

$$T(N) \otimes K \otimes T(N) \to T(N) \to T(M) \to 0,$$
$$K \otimes \mathcal{S}(N) \to \mathcal{S}(N) \to \mathcal{S}(M) \to 0,$$

and thus

$$K \otimes S(N) \to S(N) \to S(M) \to 0,$$
$$K \otimes \wedge(N) \to \wedge(N) \to \wedge(M) \to 0,$$

where the left-hand maps are given by multiplication. In particular, for every d the module $\mathcal{S}_d(M)$ *is the module* $\mathcal{S}_d(N)$ *modulo the relations* $K \cdot \mathcal{S}(N)_{d-1}$ *"generated by" K, and similarly for* $S_d(M)$ *and* $\wedge^d M$.

Remark: In part c, we have $S(M \oplus N) = S(M) \otimes S(N)$ as commutative algebras. To make $\mathcal{S}(M \oplus N) \cong \mathcal{S}(M) \otimes \mathcal{S}(N)$ and $\wedge(M \oplus N) \cong \wedge(M) \otimes \wedge(N)$ algebra isomorphisms as well, we must take the second to be the "skew tensor product of algebras": that is, $\mathcal{S}(M) \otimes \mathcal{S}(N)$ is the ordinary tensor product of graded R-modules, but the multiplication is given by the skew-commutative rule

$$a \otimes b \cdot a' \otimes b' = (-1)^{(\deg b)(\deg a')} aa' \otimes bb'$$

for homogeneous elements a, b.

Proof. All of the properties for $T(M)$ follow at once from the corresponding facts about tensor products; see Proposition A2.1. We shall deduce the statements for $\mathcal{S}$ from these (the statements for S and $\wedge$ then follow immediately, and we shall not mention them further).

a. The map φ induces a map from $T(M)$ to $T(N)$, carrying the skew-commutativity relations for $\mathcal{S}(M)$ into skew-commutativity relations for $\mathcal{S}(N)$; thus there is an induced map $\mathcal{S}(M) \to \mathcal{S}(N)$.

b. $R' \otimes_R \mathcal{S}_R(M)$ is the result of factoring out the ideal generated by elements of the form $1 \otimes (x \otimes y \pm y \otimes x)$ from $R' \otimes_R T_R(M)$. Under the identification of $R' \otimes_R T_R(M)$ with $T_{R'}(R' \otimes_R M)$, the element $1 \otimes (x \otimes y \pm y \otimes x)$ corresponds to $(1 \otimes x) \otimes (1 \otimes y) \pm (1 \otimes x) \otimes (1 \otimes y)$—and these generate the skew-commutativity relations in $T_{R'}(R' \otimes_R M)$.

c. Starting with $T(M \oplus N) = T(M) \otimes T(N) \otimes T(M) \otimes \cdots$, we first factor out the skew-commutativity relations saying that elements of M skew-commute with elements of N; the result is $T(M) \otimes T(N)$. Next we factor out the relations saying that elements of M skew-commute and elements of N skew-commute, to get $\mathcal{S}(M) \otimes_R \mathcal{S}(N)$.

d. Since the skew-commutativity relations in $T(M)$ are images of those in $T(N)$, we immediately deduce from the right-exact sequence

$$T(N) \otimes K \otimes T(N) \to T(N) \to T(M) \to 0$$

a right-exact sequence

$$\mathcal{S}(N) \otimes K \otimes \mathcal{S}(N) \to \mathcal{S}(N) \to \mathcal{S}(M) \to 0;$$

but the image of $\mathcal{S}(N) \otimes K \otimes \mathcal{S}(N)$ in $\mathcal{S}(N)$ is, by commutativity, the same as the image of $K \otimes \mathcal{S}(N)$. Since the sequence $K \otimes \mathcal{S}(N) \to \mathcal{S}(N) \to \mathcal{S}(M) \to 0$ is a direct sum of the sequences $K \otimes \mathcal{S}_{d-1}(N) \to \mathcal{S}_d(N) \to \mathcal{S}_d(M) \to 0$, the very last statement follows as well. □

A2.3.1 Bases

In the central case where M is a free module with basis $x_1, \ldots, x_r$, part c of Proposition A2.2 gives us a simple way to deduce bases of $T(M)$, $\mathcal{S}(M)$, $S(M)$, and $\wedge(M)$:

Corollary A2.3. *If M is a free module on homogeneous elements $x_1, \ldots, x_r$, then*

a. $T_d(M)$ is the free R-module of rank r^d with basis the set of all words of length d in $x_1, \ldots, x_r$.

b. $\mathcal{S}_d(M)$ is the free module with basis the set of all monomials of degree d in the x_i in which no x_i of odd degree appears to a power greater than 1.

Thus:

c. *$\mathcal{S}(M)$ is the polynomial ring on the "variables" x_i, and $\mathcal{S}_d(M)$ is the free R-module of rank $\binom{r+d-1}{r-1}$, with basis the set of monomials of degree d in the x_i.*

d. *$\wedge^d(M)$ is the free R-module of rank $\binom{r}{d}$ with basis*

$$\{x_{i_1} \wedge \cdots \wedge x_{i_d} | 1 \leq i_1 < \cdots < i_d \leq r\}$$

corresponding to the set of all d-subsets of $\{1, \ldots, r\}$.

e. *If N is another free module, with basis $y_1, \ldots, y_s$, and if $\varphi : M \to N$ is a homomorphism with matrix f, then the induced map $\wedge^d \varphi : \wedge^d M \to \wedge^d N$ has matrix whose entry corresponding to the basis elements $x_{i_1} \wedge \cdots \wedge x_{i_d}$ and $y_{j_1} \wedge \cdots \wedge y_{j_d}$ is the determinant of the submatrix of f involving the columns $i_1, \ldots, i_d$ and rows $j_1, \ldots, j_d$.*

Proof. Results a–d may be proved by induction on r, using Proposition A2.2 on a decomposition $R^r = R \oplus R^{r-1}$, and taking the grading into account in part b. Since the result is in any case fairly trivial for the tensor algebra, we illustrate with part b; Parts c and d follow from this as usual.

We may write $R^r = Rx_1 \oplus \oplus_{i=2}^r Rx_i$ as graded free modules. Now

$$\mathcal{S}(R^r) = \mathcal{S}(Rx_1) \otimes \mathcal{S}(\oplus_{i=2}^r Rx_i).$$

If degree x_1 is even, then $\mathcal{S}(Rx_1)$ has as basis the set of powers of x_1, so

$$\mathcal{S}(R^r) = (\oplus_n Rx_1^n) \otimes \mathcal{S}(\oplus_{i=2}^r Rx_i),$$

and assuming the desired result for $\mathcal{S}(\oplus_{i=2}^r Rx_i)$ by induction on r, the desired description follows by using commutativity. If, on the contrary, degree x_1 is odd, then $\mathcal{S}(Rx_1) \cong R[x_1]/(x_1)^2$, so

$$\mathcal{S}(R^r) = (R \oplus Rx_1) \otimes \mathcal{S}(\oplus_{i=2}^r Rx_i)$$
$$\mathcal{S}_d(R^r) = \mathcal{S}_d(R^{r-1}) \oplus (Rx \otimes \mathcal{S}_{d-1}(R^{r-1})),$$

and again we are done by induction and skew-commutativity.

e. To simplify the notation, we suppose that $(i_1, \ldots, i_d) = (j_1, \ldots, j_d) = (1, \ldots, d)$. We have $\wedge^d \varphi(x_1 \wedge \cdots \wedge x_d) = \varphi(x_1) \wedge \cdots \wedge \varphi(x_d)$ because $\wedge \varphi$ is an algebra map. Writing $\varphi(x_i) = \sum \varphi_{ij} y_j$, this gives

$$\wedge^d \varphi(x_1 \wedge \cdots \wedge x_d) = \sum \varphi_{1,j_1} \cdots \varphi_{d,j_d} y_{j_1} \wedge \cdots \wedge y_{j_d}.$$

Because of the skew-commutativity of $\wedge M$ we may gather the terms in which a given set of y_j appears; the coefficient of $y_1 \wedge \cdots \wedge y_d$ is

$$= \sum_\sigma \operatorname{sgn}(\sigma) \varphi_{1,\sigma(1)} \cdots \varphi_{d,\sigma(d)},$$

where σ runs over all permutations of $1, \ldots, d$, and $\operatorname{sgn}(\sigma)$ is the sign of the permutation σ, that is, $(-1)^n$, where n is the number of transpositions in representation of σ as a product of transpositions. This is just the determinant of the submatrix of φ involving rows $1, \ldots, d$ and columns $1, \ldots, d$.

□

A2.3.2 Exercises

Exercise A2.1: Show that if we define $M_{\text{even}} = \oplus_{i \text{ even}} M_i$ and $M_{\text{odd}} = \oplus_{i \text{ odd}} M_i$, then

$$\mathcal{S}(M) = S(M_{\text{even}}) \otimes \wedge(M_{\text{odd}})$$

as R-algebras.

Exercise A2.2: Compute $\wedge(M)$ and $S(M)$ when $R = \mathbf{Z}$ and $M = \mathbf{Z}/(2) \oplus \mathbf{Z}/(3)$.

Exercise A2.3:

a.* Suppose that R is a local ring and M is a finitely generated module. Compute the minimal number of generators of $S_d M$ and $\wedge^d M$ in terms of the minimal number of generators of M.

b. Suppose that R is an integral domain with quotient field $K = K(R)$. For any module M, the torsion-free rank of M is by definition the rank of the K-vector space $M \otimes_R K$. Compute the torsion-free rank of $S_d M$ and $\wedge^d M$ in terms of the torsion-free rank of M.

Exercise A2.4: Suppose that R is an integral domain, and that I is an ideal of R. Show that the natural map

$$S_d(I) \to I^d \subset R$$

has kernel equal to the torsion submodule of $S_d(I)$. Use this fact together with Exercise A2.3a to give examples of ideals I such that $S_d(I)$ has a nonzero torsion submodule. It turns out to be quite an interesting question when this and the corresponding map $S_d(I/I^2) \to I^d/I^{d+1}$ are isomorphisms; see Huneke [1982] and the references cited there for some information.

Exercise A2.5:

a. Suppose that I is an ideal in a ring R in which 2 is a unit. Show that $\wedge^d I$ is annihilated by I for every $d > 1$.

b. Under the same hypotheses as in part a, show that $\wedge^d I = \wedge^d(I/I^2)$.

Exercise A2.6: Show that if $2 = 0$ in R, then $\wedge(M) = S(M)/(\{x^2 | x \in M\})$. If M is free with basis $x_1, \ldots, x_r$, then this may also be written as

$$R[x_1, \ldots, x_r]/(x_1^2, \ldots, x_r^2).$$

Writing I for the ideal $(x_1, \ldots, x_r) \subset S := R[x_1, \ldots, x_r]$, compute the annihilator of $\wedge^2 I$. Note that it does not contain I, as would be the case if 2 were invertible. For example, if $R = k[x, y, z]$, then $xy \wedge z \neq 0 \in \wedge^2(x, y, z)$.

Exercise A2.7: The characteristic polynomial of an endomorphism has a nice expression in terms of exterior powers. Let $\varphi : F \to F$ be an endomorphism of a finitely generated free module. Show that the characteristic polynomial of φ may be expressed as

$$\det(\lambda I - \varphi) = \sum (-1)^d \text{ trace } (\wedge^d \varphi)\lambda^d.$$

Exercise A2.8 (Action of the symmetric group on a d-fold skew tensor product): If A is a graded skew-commutative algebra, then the group G of permutations of d letters acts as automorphisms of the skew-commutative algebra $A^{\otimes d}$, the d-fold skew tensor product of algebras, by permutation of the factors and multiplication by a sign. Since every permutation is the product of transpositions, the sign is determined if we specify it on transpositions; indeed, the whole action could have been defined in this way. If σ is a transposition of the ith and $(i+1)$ terms, then

$$\sigma(z_1 \otimes \cdots \otimes z_d) = (-1)^{\deg z_i \deg z_{i+1}} z_1 \otimes \cdots \otimes z_{i+1} \otimes z_i \otimes \cdots \otimes z_d.$$

Show that this really defines a representation of G as algebra automorphisms of $A^{\otimes d}$.

Exercise A2.9 (Which algebras are symmetric algebras?):* Describe the symmetric algebra of an R-module M as a quotient of a polynomial ring over R by using a free presentation of M. Use the result to characterize the quotients of polynomial rings over R that are symmetric algebras of modules.

A2.4 Coalgebra Structures and Divided Powers

In this section we shall require the following special convention for the duals of graded modules.

Notation. *If N is a graded R-module we shall write N^* for the* **graded dual** *of N, that is*

$$N^* := \oplus_d \operatorname{Hom}_R(N_d, R)$$

(rather than $N^* = \operatorname{Hom}_R(N, R)$, as we have usually written previously).

This only makes a difference when N is not a finitely generated module. It will be important for us mainly in the case $N = \mathcal{S}(M)$; then we have

$$\mathcal{S}(M)^* = \oplus_d \mathcal{S}_d(M)^*.$$

Without this convention, the dual of a graded module would not be a graded module in the usual sense: It would be the direct product, not the direct sum, of its homogeneous components.

We have seen in the previous section that $\mathcal{S}(M \oplus M) \cong \mathcal{S}(M) \otimes \mathcal{S}(M)$, the isomorphism being as skew-commutative algebras (the tensor product sign on the right is the skew tensor product; see Proposition A2.2 and the remark following it). We can exploit this by using the natural "diagonal" map $\Delta : M \to M \oplus M$ that sends x to (x, x). Δ induces a map of algebras, which we shall also call Δ,

$$\Delta : \mathcal{S}(M) \to \mathcal{S}(M \oplus M) = \mathcal{S}(M) \otimes \mathcal{S}(M),$$

and thus (by restriction) maps of algebras

$$\Delta : \wedge(M) \to \wedge(M \oplus M) = \wedge(M) \otimes \wedge(M),$$
$$\Delta : S(M) \to S(M \oplus M) = S(M) \otimes S(M),$$

all sending $m \mapsto m \otimes 1 + 1 \otimes m$ for any $m \in M$. For any graded module N we define a map $\tau : N^* \otimes N^* \to (N \otimes N)^*$ by $\tau(a \otimes b)(x \otimes y) = a(x)b(y)$ (note that there are no signs here). We dualize Δ and compose with τ to get

$$\mu = \Delta^* \tau : \mathcal{S}(M)^* \otimes \mathcal{S}(M)^* \to \mathcal{S}(M)^*.$$

The map μ is actually the multiplication map of an algebra structure on $\mathcal{S}(M)^*$; because of this Δ is called the **comultiplication** (or **diagonal**) map of $\mathcal{S}(M)$. This algebra structure on $\mathcal{S}(M)^*$ has a unit element, too, a fact which may be expressed by saying that there is a map $R \to \mathcal{S}_0(M)^*$ with certain properties. The dual of this map is called the **counit** (or **augmentation**) of $\mathcal{S}(M)$. It is defined as the projection map

$$\varepsilon : \mathcal{S}(M) \to \mathcal{S}(M)/M\mathcal{S}(M) = \mathcal{S}_0(M) = R.$$

Together, the maps Δ and ε are called the **coalgebra structure of** $\mathcal{S}(M)$. Because Δ and ε are R-algebra maps, $\mathcal{S}(M)$ is what Bourbaki calls a **bigebra**. (Others, including this author, call it a **bialgebra**, to which Bourbaki presumably objects that in arabic one would not use the article *al* with the numeral 2(*bi*).)

Another structure on $\mathcal{S}(M)$ deserves mention. It is the **antipode** map, which is the map induced on $\mathcal{S}(M)$ by the map $-1 : M \to M$. With the antipode, $\mathcal{S}(M)$ becomes what is called a **Hopf algebra**. We shall not require this structure.

Proposition A2.4. *μ gives an associative skew-commutative multiplication on $\mathcal{S}(M)^*$.*

Proof. Associativity is the statement that the two possible composites are equal in the diagram

$$(*) \qquad \mathcal{S}(M)^* \otimes \mathcal{S}(M)^* \otimes \mathcal{S}(M)^* \underset{1\otimes\mu}{\overset{\mu\otimes 1}{\rightrightarrows}} \mathcal{S}(M)^* \otimes \mathcal{S}(M)^* \xrightarrow{\mu} \mathcal{S}(M)^*.$$

It is enough to check the corresponding fact on the diagram

$$(**) \qquad \mathcal{S}(M) \xrightarrow{\Delta} \mathcal{S}(M) \otimes \mathcal{S}(M) \underset{1\otimes\Delta}{\overset{\Delta\otimes 1}{\rightrightarrows}} \mathcal{S}(M) \otimes \mathcal{S}(M) \otimes \mathcal{S}(M)$$

of which $(*)$ is the dual. As all the maps in question are algebra maps, and as $\mathcal{S}(M)$ is generated by M as an algebra, it is enough to check the desired inequality on elements of M. But identifying $\mathcal{S}(M) \otimes \mathcal{S}(M) \otimes \mathcal{S}(M)$ with $\mathcal{S}(M \oplus M \oplus M)$ by Proposition A2.2, both $(\Delta \otimes 1)\Delta$ and $(1 \otimes \Delta)\Delta$ induce the map

$$M \to M \oplus M \oplus M \qquad m \mapsto (m, m, m).$$

Similarly, skew-commutativity is the statement that the diagram in Figure A2.1 commutes, where T is the map that interchanges the two factors and introduces a sign according to our usual convention,

$$T(a \otimes b) = (-1)^{(\deg a)(\deg b)} b \otimes a \text{ for homogeneous elements } a, b,$$

and that $\mu(a \otimes a) = 0$ if $a \in \mathcal{S}(M)^*$ is homogeneous of odd degree.

(***)

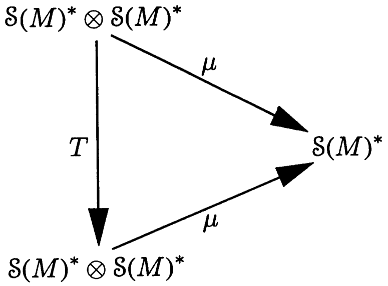


FIGURE A2.1.

Since $(* * *)$ is the dual of the diagram in Figure A2.2, we first show that the latter commutes. Because of the signs we have introduced, T is a homomorphism of algebras, and since $\mathcal{S}(M)$ is generated as an algebra by elements of M, it suffices to check that $T\Delta(m) = \Delta(m)$ for an element $m \in M$. If we identify $\mathcal{S}(M)\otimes\mathcal{S}(M)$ with $\mathcal{S}(M\oplus M)$, then $\Delta(m) = (m, m) = m \otimes 1 + 1 \otimes m$, and $T(m \otimes 1 + 1 \otimes m) = m \otimes 1 + 1 \otimes m$, as desired.

It remains to show that $\mu(n \otimes n) = 0$ for $n \in \mathcal{S}(M)^*$ homogeneous of odd degree. That is, we must show that if $a \in \mathcal{S}(M)$, then $\tau(n \otimes n)(\Delta a) = 0$. But from Lemma A2.5, below, there exists an element $b \in \mathcal{S}(M) \otimes \mathcal{S}(M)$ such that $\Delta a = b + Tb$. If we write $b = \sum b_i \otimes b_i'$, then

$$\tau(n \otimes n)(b) = \sum n(b_i')n(b_i)$$

and

$$\tau(n \otimes n)(Tb) = \sum (-1)^{\deg(b_i)\deg(b_i')} n(b_i')n(b_i).$$

Because the degree of n is odd, the only nonzero terms are those where the degrees of b_i and b_i' are both odd, so we get $n \otimes n(b) = -n \otimes n(Tb)$, whence $n \otimes n(\Delta a) = 0$ as claimed. □

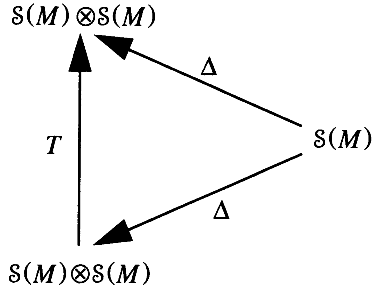


FIGURE A2.2.

Lemma A2.5 was used in the preceding proof and will be needed in a more general form shortly. Because $\mathcal{S}(M)^*$ is associative, there is a "d-fold diagonal"

$$\Delta^d : \mathcal{S}(M) \to \mathcal{S}(M)^{\otimes d} := \mathcal{S}(M) \otimes \cdots \otimes \mathcal{S}(M) \qquad (d \text{ factors}),$$

defined inductively by composing Δ^{d-1} with Δ, applied to any factor of $\mathcal{S}(M)^{\otimes(d-1)}$—the formula $(1 \otimes \Delta)\Delta = (\Delta \otimes 1)\Delta$ established in the proof of associativity in Proposition A2.4 says exactly that the choice of factor does not matter.

Lemma A2.5 (Symmetry of diagonalization). *Let G be the symmetric group on d letters, acting on $M \oplus M \oplus \cdots \oplus M$ (d times) by permuting the factors, and on*

$$\mathcal{S}(M \oplus M \oplus \cdots \oplus M) = \mathcal{S}(M)^{\otimes d}$$

by the induced maps of algebras. For any element $x \in \mathcal{S}(M)$ there is an element $y \in \mathcal{S}(M)^{\otimes d}$ such that $\Delta^d x$ is the symmetrization of y; that is,

$$\Delta^d x = \sum_{\sigma \in G} \sigma(y).$$

(*See Exercise A2.8 for another description of this action of G.*)

Proof. Since both the map Δ^d and the symmetrization map $\sum_{\sigma\in G}\sigma$ are R-linear, it suffices to prove the Lemma for elements x that are products of elements in M. We do this by induction on the number of factors. If $x \in M$ (a product with just one factor) then we check directly from the definition that $\Delta^d(x) = \sum_{\sigma\in G}\sigma(x \otimes 1 \otimes 1 \otimes \ldots \otimes 1)$, so we may take $y = x \otimes 1 \otimes 1 \otimes \ldots \otimes 1$. If x is a product of several factors we may write $x = x'x''$ in such a way that x' and x'' each involve fewer factors. By induction we may assume that $\Delta^d(x')$ and $\Delta^d(x'')$ are the symmetrizations of elements

y' and y'', respectively. Since Δ^d is an algebra homomorphism we have

$$\begin{aligned}\Delta^d(x) &= \Delta^d(x')\Delta^d(x'')\\ &= \sum_{\sigma\in G}\sigma(y')\sum_{\tau\in G}\tau(y'')\\ &= \sum_{\sigma\in G}\sum_{\tau\in G}\sigma(y'\sigma^{-1}\tau(y'')) \text{ (because } \sigma \text{ is an algebra homomorphism)}\\ &= \sum_{\sigma\in G}\sum_{\tau\in G}\sigma(y'\tau(y'')),\end{aligned}$$

the last equality holding because as τ runs over $G, \sigma^{-1}\tau$ and τ run through the same set of elements. But the last expression may be rewritten as

$$\sum_{\sigma\in G}\sigma\left(\sum_{\tau\in G}y'\tau(y'')\right);$$

that is, $\Delta^d(x)$ is the symmetrization of $\sum_{\tau\in G}(y'\tau(y''))$. This completes the induction. □

We shall now describe the structure of $\mathcal{S}(M)^*$ in the case where M is free. We shall see that if M has odd degree, so that $\mathcal{S}(M) = \wedge(M)$, we get $\wedge(M)^* \cong \wedge(M^*)$ as algebras. However, in the case when M has even degree parts, $\mathcal{S}(M)^*$ is generally a kind of algebra we have not met before, called the **divided power algebra** of M^*, We pause for definitions.

Definition. *If A is a graded skew-commutative R-algebra with $A_0 = R$, then a **system of divided powers** in A consists of a collection of functions, one for each integer $d \geq 0$:*

$$x \mapsto x^{(d)},$$

defined on the union of the A_i for $i > 0$, satisfying all the axioms that would be true if we could write $x^{(d)} = x^d/(d!)$. For elements $x, y \in \cup A_i$, these may be taken to be

$$\begin{aligned}x^{(0)} &= 1, \quad x^{(1)} = x, \quad \deg x^{(d)} = d\cdot\deg x\\ x^{(d)} &= 0 \text{ if } \deg x \text{ is odd and } d > 0\\ x^{(d)}x^{(e)} &= ((d+e)!/d!e!)x^{(d+e)}\\ (x^{(d)})^{(e)} &= ((de)!/e!(d!)^e)x^{(de)}\\ (xy)^{(d)} &= d!x^{(d)}y^{(d)} = x^d y^{(d)} = x^{(d)}y^d\\ (ax)^{(d)} &= a^d y^{(d)} \quad \text{for } a \in A_0,\end{aligned}$$

and, best of all, the "beginner's binomial theorem,"

$$(x+y)^{(d)} = \sum_{e=0}^{d} x^{(e)}y^{(d-e)}.$$

Note that from the third of these relations it follows that $d!x^{(d)} = x^d$. In an algebra over a field of characteristic 0, we may divide through by $d!$; it follows that such an algebra has at most one system of divided powers, and it is easy to check that the assignment $x^{(d)} = x^d/d!$ really does satisfy all the requirements. A more interesting example is furnished by the exterior algebra of a module; we shall do this example systematically in a moment, but we note for now that

$$(x\wedge y + z\wedge w)^2 = 2x\wedge y\wedge z\wedge w,$$

so it is natural to take $(x\wedge y + z\wedge w)^{(2)} = x\wedge y\wedge z\wedge w$, and this does extend to a system of divided powers. From this example one can see the need for the restriction that divided powers are only defined for elements of strictly positive degree: If $1^{(d)}$ were defined, then we would have $y^{(d)} = (1y)^{(d)} = d!1^{(d)}y^{(d)}$ by the next-to-last axiom. Taking $d = 2$, we get $y^{(2)} = 2 \cdot 1^{(2)}y^{(2)}$, and this would be 0 in characteristic 2. But $(x\wedge y + z\wedge w)^{(2)} \neq 0$ in the example just given, even in characteristic 2.

On the other hand, suppose that $2 = 0$ in R; in the symmetric algebra of a module of even degree, say $S = \mathcal{S}(Rx) = R[x]$, with $\deg x = 2$, we have $x^2 \neq 0$; but if S had a system of divided powers, then we would have $x^2 = 2x^{(2)} = 0$. Thus, in general, the symmetric algebra does not have divided powers.

Divided powers occur quite naturally in commutative algebra. Perhaps the most striking example is the theory of Tate, Assmus, Levin, Gulliksen, and others concerning the free resolution over a local ring (A, P) of the residue class field A/P. The general idea is to form this resolution by starting with the Koszul complex of the maximal ideal P, and "adjoin" elements to kill the cycles that are not boundaries in the Koszul complex. This adjunction may be done by exploiting the algebra structure of the Koszul complex, and adjoining new variables *with divided powers.* The ultimate result, due to Gulliksen, shows that in fact one gets the minimal free resolution this way; put more precisely, the minimal free resolution is a tensor product of exterior algebras and divided power algebras—an algebra of the form $\mathcal{S}(M)^*$ for a certain graded free module M. See the Gulliksen and Levin [1969] for an exposition.

We shall now prove the existence of divided powers in $\mathcal{S}(M)^*$. The underlying reason, as the proof will show, is the symmetric form of the diagonalization map, proved in Lemma A2.5.

Proposition–Definition A2.6. *If M is a graded free module, then the algebra structure defined above on $\mathcal{S}(M)^*$ is called the* **divided power algebra** *on M^*, and written as $\mathcal{D}(M^*)$. The algebra $\mathcal{D}(M^*)$ has a system of divided powers, which is unique, for example, if $R = \mathbf{Z}$, and may be determined by functoriality from this case.*

Proof. It is enough to prove the proposition in the case where $R = \mathbf{Z}$. As already remarked, $\mathbf{Q} \otimes \mathcal{S}(M)^*$, as an algebra containing a field of characteristic 0, trivially has a unique family of divided powers. Any system of divided powers on $\mathcal{S}(M)^*$ would pass to a system of divided powers on $\mathbf{Q} \otimes \mathcal{S}(M)^*$ by localizing. Since $\mathcal{S}(M)^*$ is a free $\mathbf{Z}$-module, nothing goes to 0 under localization, so this shows that $\mathcal{S}(M)^*$ has at most one system of divided powers. In fact we could think of the proposition as saying simply that the subalgebra

$$\mathcal{S}(M)^* = \mathbf{Z} \otimes \mathcal{S}(M)^* \subset \mathbf{Q} \otimes \mathcal{S}(M)^*$$

is closed under divided powers. (That this is not trivial, and really depends on the nature of the algebra structure map Δ^*, is shown by the fact that it would be false for $\mathcal{S}(M) \subset \mathbf{Q} \otimes \mathcal{S}(M)$.)

Let $u \in \mathcal{S}_n(M)^*$; we must prove that u^d is divisible by $d!$. If n is odd then $u^d = 0$, so we may assume that n is even. Now multiplication in $\mathcal{S}(M)^*$ is defined as the dual of the diagonal on $\mathcal{S}(M)$; this means that as a functional on $\mathcal{S}(M)$, the value of u^d on an element $x \in \mathcal{S}(M)$ is the value of $u^{\otimes d} \in \mathcal{S}(M)^{*\otimes d}$ on $\Delta^d(x) \in \mathcal{S}(M)^{\otimes d}$. By Lemma A2.5, $\Delta^d(x)$ is the "symmetrization" of some element $y \in \mathcal{S}(M)^{\otimes d}$; that is, we may write $\Delta^d(x) = \sum_{\sigma \in G} \sigma(y)$. But $u^{\otimes d}$ acts nontrivially only on tensor products of d elements of degree n, and since n is even, σ acts on such a tensor product by permutation of the factors alone—that is, with a sign of $+1$. Thus, $u^{\otimes d}(y) = u^{\otimes d}(\sigma(y))$ for any σ, so $u^{\otimes d}(x) = d! u^{\otimes d}(y)$, as required.

If M is free of finite rank, then $\mathcal{S}(M)^*$ is a bialgebra too, with dual $\mathcal{S}(M)$: In this case the natural map $\mathcal{S}(M)^* \otimes \mathcal{S}(M)^* \to (\mathcal{S}(M) \otimes \mathcal{S}(M))^*$ is an isomorphism, so we may define the diagonal map (written, as usual, as Δ) on $\mathcal{S}(M)^*$ to be the dual of the multiplication map on $\mathcal{S}(M)$. This map Δ is an algebra homomorphism as before, and it is easy to check that if $\alpha \in M^* = \mathcal{S}_1(M)^*$, then $\Delta(\alpha) = \alpha \otimes 1 + 1 \otimes \alpha$.

In the case where M is free we can give an explicit basis for $\mathcal{S}(M)^*$, dual to the monomial basis for $\mathcal{S}(M)$. The following proposition is important because it describes the multiplication in terms of the basis.

Proposition A2.7.

a. *If M has free generators $x_1, \ldots, x_m$ in odd degrees and $y_1, \ldots, y_n$ in even degrees, and if we write ξ_i and η_i for the dual basis elements of x_i and y_i, then $\mathcal{D}(M^*)$ is freely generated as a module by "divided monomials," where a divided monomial is something of the form*

$$\xi_{i_1} \wedge \cdots \wedge \xi_{i_s} \cdot \eta_1^{(d_1)} \cdots \eta_n^{(d_n)} \quad \textit{with } i_1 < \cdots < i_s \le m.$$

In fact, the divided monomials form (up to sign) a dual basis to the basis of monomials

$$x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}$$

of $\mathcal{S}(M)$.

Thus, writing $\mathcal{D}(\alpha)$ *for the divided power algebra of a free module with one generator* α, *we have*

$$\mathcal{D}(M^*) \cong \mathcal{D}(\xi_1) \otimes \cdots \otimes \mathcal{D}(\xi_s) \otimes \mathcal{D}(\eta_1) \otimes \cdots \otimes \mathcal{D}(\eta_n).$$

b. *If* M *is a free module, then* $\wedge(M)^* \cong \wedge(M^*)$ *as algebras.*

c. *If* M *is a free module and* R *contains the rational numbers, then* $\mathcal{D}(M^*) = \mathcal{S}(M)^* \cong \mathcal{S}(M^*)$ *as algebras.*

Proof.

a. It suffices to check that $\xi_{i_1} \wedge \cdots \wedge \xi_{i_s} \cdot \eta_1^{(d_1)} \cdots \eta_n^{(d_n)}$ is the dual basis element to $x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}$, and for this it suffices to treat the case of free modules over **Z** or even over **Q**. Setting $d = s + \sum d_i$, we have (for any element x of M) $\Delta^d(x) = x \otimes 1 \otimes \cdots \otimes 1 + 1 \otimes x \otimes 1 \otimes \cdots \otimes 1 + \cdots$, so

$$\Delta^d(x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}) = \prod \Delta(x_{i_1}) \prod \Delta(y_i)^{d_i}$$

is a sum of terms each of which comes from a permutation of the factors of $x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}$ and a separation of these factors into d groups. Now $\xi_{i_1} \wedge \cdots \wedge \xi_{i_s} \cdot \eta_1^{(d_1)} \cdots \eta_n^{(d_n)}$ acts on a monomial ζ such as $x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}$ by the action of

$$(1/d_1!d_2! \cdots d_n!)\xi_{i_1} \otimes \cdots \otimes \xi_{i_s} \otimes \eta_1^{\otimes d_1} \cdots \eta_n^{\otimes d_n}$$

on $\Delta^d(\zeta)$. Thus the value can be nonzero only when $\zeta = x_{i_1} \wedge \cdots \wedge x_{i_s} \cdot y_1^{d_1} \cdots y_n^{d_n}$, and it is ± 1 in this case by inspection.

b. For any module M, we have seen in Proposition A2.4 that $\wedge(M)^*$ is an algebra satisfying the same kind of skew-commutativity properties as $\wedge(M^*)$; thus there is a natural map of algebras $\wedge(M^*) \to \wedge(M)^*$ induced by the identity map $M^* \to M^*$. In the case of a free module M, we have seen from part a that the basis of $\wedge(M)^*$ corresponds to the basis of $\wedge(M^*)$ under this map, which is thus an isomorphism. The proof of part c is similar. □

A2.4.1 $\mathcal{S}(M)^*$ *and* $\mathcal{S}(M)$ *as Modules over One Another*

Notation: If $a \in A$ and $\alpha \in A^*$, then to avoid confusion with the following definitions we shall write $\langle a, \alpha \rangle \in R$ for the result of applying α to a, so that $\langle , \rangle : A \otimes A^* \to R$ is the natural pairing.

If A is any R-algebra, then $A^* = \mathrm{Hom}_R(A, R)$ is naturally an A-module: If $a \in A$ and $\alpha \in A^*$, then we define $a(\alpha) \in A^*$ by means of the formula $(a(\alpha))(b) = \langle ab, \alpha\rangle$. The reason for the notation $a(\alpha)$, in place of the more ordinary looking $a\alpha$, is that we shall shortly be dealing with the case where A is a bialgebra and $A = A^{**}$, so we must distinguish between $a(\alpha) \in A^*$ and $\alpha(a) \in A$. The module structure has a nice description in terms of the diagonal of A^* (which is by definition the dual of the multiplication map $A \otimes A \to A$): It is given by the composition

$$A \otimes A^* \xrightarrow{1\otimes\Delta} A \otimes A^* \otimes A^* \xrightarrow{\langle,\rangle\otimes 1} R \otimes A^* = A^*,$$

as the reader may easily check. In more concrete terms, if we write $\Delta\alpha$ as $\sum \alpha_i \otimes \alpha_i'$, then $a(\alpha) = \sum\langle a, \alpha_i\rangle\alpha_i'$.

If A is a bialgebra like $\mathcal{S}(M)$ or (in case M is free) $\mathcal{S}(M)^*$, then it makes sense to ask how this module structure interacts with the algebra structure. The following result gives a piece of the answer.

Proposition A2.8. *If $\alpha \in M^*$, then α acts as a derivation on $\mathcal{S}(M)$ in the sense that if $a, b \in \mathcal{S}(M)$ are homogeneous, then*

$$\alpha(ab) = \alpha(a)b + (-1)^{(\deg\alpha)(\deg a)}a\alpha(b).$$

Thus, for example, if M is free and R contains a field of characteristic 0, $S(M)^*$ may be identified as the algebra of differential operators with constant coefficients on $S(M)$.

Proof. We need the following observation about the diagonal: For $a \in M$, the diagonal is $\Delta a = a \otimes 1 + 1 \otimes a$. Further, since $\mathcal{S}(M)$ is a bialgebra, $\Delta(ab) = (\Delta a)(\Delta b)$. Since any element of $\mathcal{S}(M)$ is a sum of products of elements of M, we see that for any $a \in \mathcal{S}(M)_+ := \sum_{j>0}\mathcal{S}_j(M)$ we may write

$$\Delta a = 1 \otimes a + a \otimes 1 + \sum m_i \otimes m_i'$$

where $m_i, m_i' \in \mathcal{S}(M)_+$. Returning to the proof of the proposition, we write $\Delta a = \sum a_i \otimes a_i'$ and $\Delta b = \sum b_i \otimes b_i'$. We have

$$\Delta(ab) = \sum_{i,j}(-1)^{(\deg a_i')(\deg b_j)}a_ib_j \otimes a_i'b_j',$$

so

$$\alpha(ab) = \sum_{i,j}(-1)^{(\deg a_i')(\deg b_j)}\langle\alpha, a_ib_j\rangle a_i'b_j'.$$

Since $\alpha \in M^*$, the only nonzero terms are those for which $a_ib_j \in M = \mathcal{S}_1M$, that is, where one of a_i, b_j is in $\mathcal{S}_1M$ and the other is in $\mathcal{S}_0M$. By the remark on the form of Δ above, we get

$$\begin{aligned}\alpha(ab) &= \sum_i \langle \alpha, a_i \rangle a_i' b + \sum_j (-1)^{(\deg a)} \langle \alpha, b_j \rangle a b_j' \\ &= \alpha(a) b + (-1)^{\deg a} a \alpha(b),\end{aligned}$$

as required.

Some consequences of this for the exterior algebra are described near the end of Chapter 17.

A2.5 Schur Functors

In the introduction to this appendix we mentioned the connection of multilinear algebra and representation theory. Here we sketch the beginnings of it. We have included this material simply for its intrinsic interest; it is not used elsewhere in this book.

Let V be a finite-dimensional vector space over a field k. Let $G = \mathrm{SL}(V)$ be the group of linear transformations of V with determinant 1. The group G acts on $T_d V = V^{\otimes d}$ by acting on each factor. For $\sigma \in G$ and $v_1 \otimes \cdots \otimes v_d \in V^{\otimes d}$, we set

$$\sigma(v_1 \otimes \cdots \otimes v_d) = \sigma v_1 \otimes \cdots \otimes \sigma v_d.$$

It turns out that if k has characteristic 0, then every representation of G on a finite-dimensional k-vector space is a direct sum of irreducible representations (one says that G is "reductive"). Further more, every irreducible representation occurs as a summand of some $T_d V$. These summands are also irreducible representations of $\mathrm{GL}(V)$. To get all irreducible representations of $\mathrm{GL}(V)$, one simply tensors these irreducible representations with 1-dimensional representations of the form

$$\sigma(r) = (\det \sigma)^m r \quad \text{for } r \in k,$$

for $m \in \mathbf{Z}$.

One may think of the process of making a representation as the process of constructing a new vector space from V in some reasonably functorial way. For example, $\wedge^m V$ and $S_m V$, the exterior and symmetric powers are representations of $\mathrm{SL}(V)$, or even of $\mathrm{GL}(V)$. But $\wedge^m$ and S_m have a stronger property: They are functors in the sense that any linear transformation of V acts naturally on $\wedge^m V$ and $S_m V$. As we have seen, $\wedge^m$ and S_m may even be regarded as functors on any category of modules. The same thing turns out to be true for all the irreducible representations mentioned above, and some others as well. The resulting functors, which include $\wedge^m$ and S_m, are called **Schur functors** in honor of Issai Schur, one of the pioneers of representation theory in the beginning of this century.

To give the definition, let R be any ring and let M be any R-module. Let A be a finite set and let $\mathcal{A}$ be a pair of partitions $\{A_i\}_{i=1\cdots s}$ and $\{B_j\}_{j=1\cdots t}$ of A into disjoint subsets: That is, writing $\amalg$ for disjoint union, we have

$$\mathcal{A} : A = \amalg_{i=1,\ldots,s} A_i = \amalg_{j=1,\ldots,t} B_j.$$

For any integer d, the diagonal Δ of the exterior algebra on M gives us a map $\Delta^d : \wedge^d M \to T_d M = M^{\otimes d}$. If we think of a collection of copies of M indexed by the elements of A, and if $A_i \subset A$ is a subset of cardinality d, then we shall write $\wedge^{A_i} M \to T_{A_i} M$ for this map $\wedge^d M \to T_d M$, thought of as involving the copies of M labeled by the elements of A_i. By tensoring several such maps together we get a map

$$\alpha : \wedge^{A_1} M \otimes \cdots \otimes \wedge^{A_s} M \to T_{A_1} M \otimes \cdots \otimes T_{A_s} M.$$

There is a natural identification

$$T_{A_1} M \otimes \cdots \otimes T_{A_s} M = T_{B_1} M \otimes \cdots \otimes T_{B_t} M$$

obtained by using the associativity and commutativity of the tensor product. For any integer d the multiplication map in the symmetric algebra of M gives a map $T_d M \to S_d M$. As above, if B_j is a subset of A of cardinality d, we write $T_{B_j} M \to S_{B_j} M$ for $T_d M \to S_d M$, thought of as involving the copies of M indexed by elements of B_j. By tensoring several such maps together we get a map

$$\beta : T_{B_1} M \otimes \cdots \otimes T_{B_t} M \to S_{B_1} M \otimes \cdots \otimes S_{B_t} M.$$

We define the Schur module $S^{\mathcal{A}} M$, associated to $\mathcal{A}$ and M, to be the image of the composite map

$$\beta\alpha : \wedge^{A_1} M \otimes \cdots \otimes \wedge^{A_s} M \to S_{B_1} M \otimes \cdots \otimes S_{B_t} M.$$

This construction is obviously functorial in M (that is, if $M \to N$ is any homomorphism, then there is an induced homomorphism $S^{\mathcal{A}} M \to S^{\mathcal{A}} N$, with compositions preserved); the functor $S^{\mathcal{A}}$ is the Schur functor associated to the pair of partitions.

It should be said that this definition is one of several possible definitions which are equivalent in the central case where R is a field and M is a finitely generated vector space. It is known to have reasonable properties when R is any ring and M is a free module—for example, $S^{\mathcal{A}} M$ is again a free module, whose rank depends only on the rank of M and the partitions $\mathcal{A}$ of A, not on the properties of R. However, the case where M is not free has hardly been investigated, and the definition may be the "wrong" one in this generality. For some other possibilities see Towber [1979] and Akin, Buchsbaum, and Weyman [1982].

First examples for the reader to check: If all the A_i have cardinality 1 and $B_1 = \mathcal{A}$ has cardinality d, then the corresponding Schur functor $S^{\mathcal{A}} M = S_d M$ is a symmetric power. If, on the other hand, the module M is free, all the B_i have cardinality 1, and $A_1 = \mathcal{A}$, then $S^{\mathcal{A}} M = \wedge^d M$. (See Exercise A2.14 for the kind of trouble that can result if M is not free.)

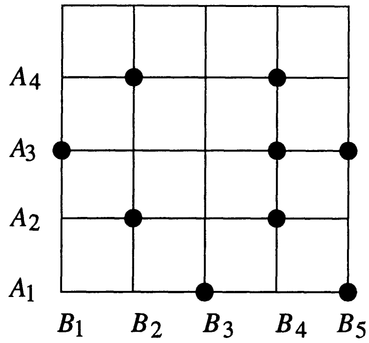


FIGURE A2.3.

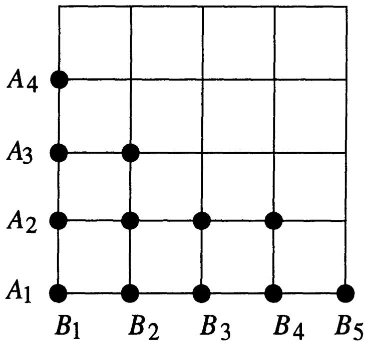


FIGURE A2.4.

It is easy to see that if any B_j contains more than one element of some A_i, then the corresponding Schur functor is 0. Thus the only interesting pairs of partitions are those constructed as follows: Let A be a set of lattice points in the plane, and let $\mathcal{A}$ be defined by taking A_i to be the set of points in the ith row of A, while B_j is the set of points in the jth column. An example is shown in Figure A2.3, where A is the set consisting of the 9 heavy dots. If R is a field of characteristic 0, then the irreducible representations of $\mathrm{SL}(V)$ are given by those Schur functors $S^{\mathcal{A}}V$ that correspond to the sets of lattice points of the form shown in Figure A2.4, with "width" less than $\dim V$.

More precisely, given a sequence of integers $d_1 \geq d_2 \geq \cdots \geq d_s \geq 0$, called a **partition** of $d = \sum d_i$, we define a set A of d lattice points in the plane as follows: Put d_i lattice points consecutively in the ith row (positions $(i,1)$–(i,d_i)). The pair of partitions of A by rows and columns then partitions A into sets A_i of cardinality d_i and B_j of cardinality e_j. One usually writes $S^{\{d_1,\ldots,d_s\}} = S^{\mathcal{A}}$ for the corresponding Schur functor. Thus, if $\{1^d\}$ represents a sequence of ones $\{1,1,\ldots,1\}$ of length d, then $S^{\{1^d\}}M$ is the dth symmetric power of M, while $S^{\{d\}}M$ is the dth exterior power

$$\begin{vmatrix} \binom{r}{d_1} & \binom{r}{d_2-1} & & & \\ \binom{r}{d_1+1} & \binom{r}{d_2} & & & \\ & & \begin{matrix} \vdots \\ \cdots\cdots \\ \vdots \end{matrix} & & \\ & & & & \binom{r}{d_s-1} \\ & & & \binom{r}{d_1+1} & \binom{r}{d_s} \end{vmatrix}$$

FIGURE A2.5.

of M. The first "new" object is $S^{\{2,2\}}M$. We shall not prove the following theorem.

Theorem A2.9. *If M is a free R-module of rank r, and all the d_i are less than r, then $S^{\{d_1,\ldots,d_s\}}M$ is a free R-module of rank equal to the determinant of binomial coefficients shown in Figure A2.5. If R is a field of characteristic 0, and V is a vector space of dimension r, then the $S^{\{d_1,\ldots,d_s\}}V$ with $r > d_1 \geq d_2 \geq \cdots \geq d_s > 0$ are the distinct nontrivial, irreducible, finite-dimensional representations of* $\mathrm{SL}(V)$.

For example, $S^{\{2,2\}}M$ is the image of $\wedge^2 M \otimes \wedge^2 M$ in $S_2 M \otimes S_2 M$ under the map sending $a\wedge b \otimes c \wedge d$ to $ac\otimes bd - bc\otimes ad - ad\otimes bc + bd\otimes ac$. The kernel of this map is (in characteristic 0) in fact $\wedge^3 M \otimes M$, embedded in $\wedge^2 M \otimes \wedge^2 M$ by the map diagonalizing the $\wedge^3 M$ into $\wedge^2 M \otimes M$, and then wedging the last factors; that is, if $m\otimes n \in \wedge^3 M \otimes M$ and $\Delta m = \sum m_i \otimes m_i' \in \wedge^2 M \otimes M$, then $m \otimes n \mapsto \sum m_i \otimes (m_i' \wedge n) \in \wedge^2 M \otimes \wedge^2 M$. If M has rank r, then $S^{\{2,2\}}M$ is a free module of rank $\binom{r}{2}^2 - \binom{r}{1}\binom{r}{3}$.

For a further treatment of these matters, see Fulton and Harris [1992].

A2.5.1 Exercises

Exercise A2.10 (Divided powers and the rational normal curve): The rational normal curve is another construction that shows how "right" the divided powers are. We begin with a reminder of the classical situation:

a. Let k be a field of characteristic 0, and let V be a k-vector space of dimension 2 with basis s, t. The subset of $S_d(V)$ consisting of dth powers of linear forms is the affine cone over a projective curve, the rational normal curve in $\mathbf{P}^d = \mathbf{P}(S_d V)$. If we take the coefficient of the monomial $s^i t^{d-i}$ to be the variable x_i, then the homogeneous ideal of the curve is generated by the 2×2 minors of the matrix

$$\begin{pmatrix} (d)x_0 & (d-1)x_1 & \cdots & (1)x_{d-1} \\ (1)x_1 & (2)x_2 & \cdots & (d)x_d \end{pmatrix}.$$

Show at least that the 2×2 minors of this matrix do vanish when we substitute for the x_i the coefficients of the dth power of a linear form.

b. Show that if d is a power of a prime p, and k is a field of characteristic p, then the dth powers of linear forms in $k[s,t]$ form a 2-dimensional linear subspace of $S_d(V)$—not the cone over a rational normal curve of degree d.

c. Independently of characteristic, the divided dth powers of linear forms do sweep out the rational normal curve, and the homogeneous ideal of the curve is generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x_0 & x_1 & \ldots & x_{d-1} \\ x_1 & x_2 & \ldots & x_d \end{pmatrix}.$$

Show at least that the 2×2 minors of this matrix do vanish when we substitute for the x_i the coefficients of the dth divided power of a linear form.

Exercise A2.11 (Divided powers and Pfaffians): Let F be a free module and let $\varphi : F^* \to F$ be a homomorphism. We say that φ is **alternating** if φ is skew-symmetric (that is, $\varphi^* = -\varphi$) and $x(\varphi(x)) = 0$ for every $x \in F^*$. (Note that if 2 is a unit in R, then each of these two conditions implies the other.)

a. Show that there is a one-to-one correspondence between elements of $\wedge^2 F$ and alternating homomorphisms $F^* \to F$ given by sending $\alpha \in \wedge^2 F$ to the map $\varphi_\alpha : x \mapsto x(\alpha) \in F$.

b. If φ is any alternating map, then the determinant of φ is well defined up to a square in R as the determinant of φ with respect to any basis of F and its dual basis in F^*. If rank F is odd, then $\det \varphi = 0$.

c.* Suppose rank $F = 2r$ is even, and consider an alternating homomorphism $\varphi : F^* \to F$ and the corresponding element $\alpha \in \wedge^2 F$. If we choose a basis e_i for F, we may write the rth divided power of α as

$$\alpha^{(r)} = u e_1 \wedge \cdots \wedge e_{2r} \quad \text{for some } u \in R.$$

The element u is called the **Pfaffian** of φ, written $u = \text{Pfaff}(\varphi)$. Show that if we compute with respect to the basis e_i of F and a dual basis of F^*, then

$$\det \varphi = (\text{Pfaff } \varphi)^2.$$

d. More generally, show that the coefficient in $\alpha^{(s)}$ of a basis element $e_{i_1} \wedge \cdots \wedge e_{i_{2s}}$ of $\wedge^{2s} F$ is the Pfaffian of the alternating submatrix of φ obtained by taking rows and columns numbered $i_1, \ldots, i_{2s}$.

Exercise A2.12: If M is a free module, then the dual of the multiplication map of $\mathcal{S}(M)$ is a comultiplication for $\mathcal{D}(M^*)$. Show that $\mathcal{D}(M^*)$ is a bialgebra too.

Exercise A2.13:* Let k be a field, let $R = k[[x,y]]/(x^2 - y^3)$, and let $I = (x,y)$ be the maximal ideal. Show that a free presentation of I is given by the matrix

$$\begin{pmatrix} y & x \\ -x & y^2 \end{pmatrix}.$$

Deduce:

a. $I^* \cong I$

b. $\wedge^2 I = k$

c. $S_2 I$ requires three generators.

d. $(S_2 I)^*$ requires only two generators.

Conclude that $\wedge(I)^*$ is not the exterior algebra of any module, and that $S(I)^*$ is not the symmetric algebra of any module, even in characteristic 0.

Exercise A2.14: Consider the part of the diagonal map that goes from $\mathcal{S}_d M \to M \otimes \cdots \otimes M$ (d factors); call it ι.

a. Show that if M is a free module, then ι is an injection.

b. Show that if R contains a field of characteristic 0, then ι is an injection.

c. Show that if $R = k[x,y,z]$, $M = (x,y,z)$, and k is a field of characteristic 2, then $\iota : \wedge^2 M \to M \otimes M$ is not an injection; in fact show that $\iota(xy \wedge z) = 0$, and use the result of Exercise A2.6.

A2.6 Complexes Constructed by Multilinear Algebra

Certain complexes that are useful in commutative algebra and algebraic geometry can be defined easily in terms of the multilinear constructions we have already made. In this section we sketch the construction and basic properties of a family of complexes beginning with the Eagon-Northcott complex and the Buchsbaum-Rim complex.

The Koszul complex is associated to a sequence of elements of a ring R, or equivalently to a map from a finitely generated free R-module F to R. The complexes defined in this section are associated in an analogous way to an arbitrary map of finitely generated free modules $F \to G$. (The definitions and results that follow can be extended easily to the case where F and G are only locally free; we remark on the necessary modifications in the text.) But even though they generalize the Koszul complex, they are also contained in the Koszul complex in a certain sense.

Let $\varphi : F \to G$ be a map of free modules over a ring R. Write f for the rank of F and g for the rank of G. Although much of what we shall do works formally for any f, g, the applications will all be in the case where $f \geq g$, so we shall assume from now on that $f \geq g$ and think of φ as a presentation of the R-module $M := \operatorname{coker} \varphi$. By Fitting's lemma (Corollary 20.4), M is annihilated by the ideal $I_g(\varphi)$ generated by the $g \times g$ minors of φ, as are a whole host of associated modules such as the symmetric powers of M. We shall see that under reasonable assumptions M and these other modules are isomorphic to ideals of codimension 1 in $R/I_g(\varphi)$, and they arise naturally in the divisor theory of determinantal varieties. We shall define complexes $\mathcal{C}^i$ that should be thought of as attempts at resolutions for some of these modules. The complexes $\mathcal{C}^i$ for $i \geq -1$ will actually be resolutions in the "generic" case, where φ is represented by a matrix of indeterminates $r_{i,j}$ over a polynomial ring $\mathbf{Z}[\{r_{i,j}\}_{i=1\cdots g, j=1\cdots f}]$.

For example, $\mathcal{C}^0$, the Eagon-Northcott complex, provides a resolution for $R/I_g(\varphi)$ in the generic case, and $\mathcal{C}^1$, the Buchsbaum-Rim complex, does the same for M itself. These two complexes are resolutions whenever $I_g(\varphi)$ contains a regular sequence of length $f - g + 1$. The conditions for some of the other complexes in the family to be resolutions involve lower order minors as well.

The motivation behind the discovery of the Eagon-Northcott complex was to understand Macaulay's unmixedness theorem (Corollary 18.14), which was strengthened, in Eagon's thesis, to say that the ring $R/I_k(\varphi)$ is Cohen-Macaulay for generic φ and all $0 \leq k \leq g$. By the Auslander-Buchsbaum formula, this is equivalent to the existence of a free resolution of $R/I_k(\varphi)$ of length $(f-k+1)(g-k+1)$; Buchsbaum had suggested finding a proof along these lines. In the case $k = g$, Eagon and Northcott accomplished this by defining the Eagon-Northcott complex [1962], and Buchsbaum [1964] gave a different, nonminimal resolution. The Buchsbaum-Rim complex was defined in a subsequent paper (Buchsbaum-Rim [1964]).

Buchsbaum and Eisenbud [1973] noted that the Eagon-Northcott and Buchsbaum-Rim complexes fit into the family of complexes presented here, and our treatment follows ideas sketched in that paper. The same family was discovered independently by David Kirby [1974]. A geometric idea originating in George Kempf's (unpublished) thesis [1970] represents the same complexes as pushforwards of Koszul complexes defined on Grassmann bundles, computed using Bott's vanishing theorem and its relatives.

This technique generalizes, in principle, to many other complexes, including those giving resolutions of lower order minors. See Weyman [1990] for an exposition and a beautiful recent application.

Many mathematicians, starting perhaps with Buchsbaum [1970], have tried to find complexes related to the lower order minors of φ in the same way that the Eagon-Northcott complex or other members of the families here are related to the maximal minors. Through work of Lascoux [1978], who used the idea of Kempf mentioned above; Nielsen [1981]; Akin, Buchsbaum, and Weyman [1982]; Pragacz and Weyman [1985] and others, the situation is reasonably well understood in characteristic 0, but many questions about the form of these complexes remain unanswered in more general cases.

For a spectacular geometric application of the Eagon-Northcott complex (in a case where it is not even exact!), see Gruson, Lazarsfeld, and Peskine [1983]. Buchsbaum and Rim were motivated by a desire to generalize the multiplicity theory that comes from the Koszul complex. The "Buchbaum-Rim" multiplicity that they defined has recently found interesting geometric applications in the work of Kleiman, Thorup, Gaffney, Rees, and others (see, for example, Kleiman and Thorup [in press]). Geometric applications of the family of complexes defined here have been made by Schreyer [1986] and others.

A2.6.1 Strands of the Koszul Complex

We begin with a reinterpretation of the Koszul complex. Let $S = S(G)$, the symmetric algebra on G (that is, the graded polynomial ring on a set of free generators for G, regarded as elements of degree 1). If $x_1, \ldots, x_g$ is a free basis of G, then $S = R[x_1, \ldots, x_g]$. If we write F' for the S-module $S \otimes F(-1)$, then there is a unique map of S-modules $\varphi' : F' \to S$ that sends $F = R \otimes F = (S \otimes F)_0 = F'_1 \subset F'$ to $G = S_1 G$ by φ. In terms of bases, let $e_1, \ldots, e_f$ be free generators of F, and suppose $\varphi e_i = \Sigma r_{i,j} x_j$ with $r_{i,j} \in R$. If we write e'_i for the generator $1 \otimes e_i \in F'$, then $\varphi' e'_i = \Sigma r_{i,j} x_j$, thought of as an element of $S = R[x_1, \ldots, x_g]$.

Let

$$K(\varphi') : 0 \to \wedge^f F' \to \wedge^{f-1} F' \to \cdots \to \wedge^2 F' \to F' \xrightarrow{\varphi'} S$$

be the Koszul complex determined by φ' over the ring S. Since the boundary maps of $K(\varphi')$ are homogeneous of degree 0 in the sense of the grading on S, we may restrict to a single degree and derive a complex of free R-modules, called a **strand** of $K(\varphi')$. Explicitly, in degree d we have a complex

$$K(\varphi')_d : \ \cdots \xrightarrow{\partial} S_{d-i}G \otimes \wedge^i F \xrightarrow{\partial} S_{d-i+1}G \otimes \wedge^{i-1} F$$
$$\xrightarrow{\partial} \cdots \xrightarrow{\partial} S_{d-1}G \otimes F \xrightarrow{\partial} S_d G,$$

where for simplicity we call each of the differentials ∂. If we write $\{\hat{x}_i\}$ for the basis of G^* dual to the basis $\{x_i\}$ of G, then $\varphi^*(\hat{x}_i) \in F^*$ acts on $\wedge F$,

and the map ∂ takes an element $m \otimes f \in S_{d-i}G \otimes \wedge^i F$ to the element $\Sigma_i x_i m \otimes \varphi^*(\hat{x}_i)(f) \in S_{d-i+1}G \otimes \wedge^{i-1}F$.

The element $c = \Sigma x_i \otimes \hat{x}_i \in G \otimes G^*$, sometimes called the "trace" element, is independent of the basis x_i chosen; it is the image of the identity element under the map $R \to G \otimes G^*$ that is dual to the evaluation map $G^* \otimes G \to R$. The maps ∂ are simply given by multiplication by c, regarded as an element of $S(G) \otimes \wedge G^*$ (where the action of $\wedge G^*$ on $\wedge F$ is via the algebra map $\wedge\varphi^* : \wedge G^* \to \wedge F^*$). In particular, since $S(G) \otimes \wedge G^*$ is the exterior algebra over SG of the SG-module $SG \otimes G^*$, and c is an element of degree 1 in this exterior algebra, we see at once that $c^2 = 0$ —another proof that $\partial^2 = 0$.

Now dualizing $K(\varphi')_d$ by taking $\mathrm{Hom}(-, R)$, we derive a complex

$$K(\varphi')_d^* : \cdots \to (S_{d-i+1}G)^* \otimes \wedge^{i-1}F^* \to (S_{d-i}G)^* \otimes \wedge^i F^* \to \cdots$$

If we choose an element $\alpha \in \wedge^f F$ then we may use it to identify $\wedge^i F^*$ with $\wedge^{f-i}F$ as in Chapter 17.[2] To simplify the notation we shall also write D_iG^* (which we may think of as a component of the divided power algebra of G^*) in place of $(S_iG)^*$. We may now rewrite this dual complex as

$$\begin{aligned} K(\varphi')_d^* :\ 0 &\longrightarrow D_dG^* \otimes \wedge^f F \xrightarrow{\delta} D_{d-1}G^* \otimes \wedge^{f-1}F \xrightarrow{\delta} \cdots \\ &\xrightarrow{\delta} D_{d-i}G^* \otimes \wedge^{f-i}F \xrightarrow{\delta} D_{d-i-1}G^* \otimes \wedge^{f-i-1}F \xrightarrow{\delta} \cdots . \end{aligned}$$

Since DG^* is a module over SG, the map δ may again be described as multiplication by $c \in G \otimes G^* \subset SG \otimes \wedge G^*$. Note that the entries in a matrix for any ∂ or δ will be R-linear forms in the entries of a matrix for φ.

The complexes we are interested in are the $K(\varphi')_d$, the $K(\varphi')_d^*$, and, most important, a type made by "splicing together" $K(\varphi')_d^*$ with $K(\varphi')_{f-g-d}$, in case $d \leq f - g$. Note that if $d \leq f - g$, then the last term of $K(\varphi')_d^*$ at the right end is $D_0G^* \otimes \wedge^{f-d}F$, which we may identify with $\wedge^{f-d}F$, since $D_0G^* = R$. We shall define a "splice" map ε from $\wedge^{f-d}F$ to the first term at the left end of $K(\varphi')_{f-g-d}$, which is $\wedge^{f-g-d}F$. We shall denote by $\mathcal{C}^i$ the complex obtained by splicing $K(\varphi')_{f-g-i}^*$ and $K(\varphi')_i$.

We shall define ε in terms of a generator $\gamma \in \wedge^g G^* \cong R$; such a γ may be chosen because we have assumed that G is free.[3] Given the choice of γ, we define, for any $k \geq g$, the map

$$\varepsilon : \wedge^k F \to \wedge^{k-g}F$$

to be the action of $\wedge^g \varphi^* \gamma$ on F. In terms of bases, if we write I for a subset $i_1 < \cdots < i_k$ of $\{1, \ldots, f\}$, and $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$ for the corresponding product, then

---

[2]A choice of generator for $\wedge^f F$ is possible in our case because we have assumed that F is free. If we had only assumed that F was locally free, then we would have had to tensor the whole complex with $\wedge^f F$ at this point.

[3]If we had only assumed that G was locally free, then we would have had to tensor part of the complex with $\wedge^g G$ to be able to make the definitions properly.

$$\varepsilon(e_I) = \sum_{J\subset I, |J|=g} \operatorname{sgn}(J \subset I)(\det \varphi_J)e_{I-J},$$

where φ_J is the $g \times g$ submatrix of φ with columns corresponding to the basis elements indexed by J, $\operatorname{sgn}(J \subset I)$ is the sign of the permutation of I that puts the elements of J into the first g positions, and e_{I-J} is the wedge product of the basis vectors indexed by elements of the set $I - J$. The entries of a matrix for ε are thus $g \times g$ minors of φ—in particular, they are forms of degree g in the entries of a matrix for φ. For example, in case $k = g$,

$$\varepsilon : \wedge^g F \to \wedge^0 F = R$$

may be identified with the composite

$$\wedge^g F \xrightarrow{\wedge^g \varphi} \wedge^g G \overset{\gamma}{\cong} R,$$

whose image is the ideal of $g \times g$ minors of φ.

We are now ready to write down the complexes. To simplify the notation, we shall write

S_d for $S_d G$,

D_d for $D_d G^* = (S_d G)^*$,

$\wedge^d$ for $\wedge^d F$.

We also note that $D_0 = S_0 = \wedge^0 = R$, and we shall substitute R for these wherever they occur; in particular, we shall allow ourselves to think of $\varepsilon : \wedge^k F \to \wedge^{k-g} F$ as a map from $D_0 G^* \otimes \wedge^k F$ to $\wedge^{k-g} F$. Further, we write G, G^*, and F in place of $D_1 = D_1 G^*, S_1 = S_1 G$, and $\wedge^1 = \wedge^1 F$.

A description of two leading cases will serve to illustrate these conventions. Perhaps the most important case is that of the Eagon-Northcott complex

$$\begin{aligned}\mathcal{C}^0 : 0 \longrightarrow D_{f-g}G^* \otimes \wedge^f F &\xrightarrow{\delta} D_{f-g-1}G^* \otimes \wedge^{f-1} F \xrightarrow{\delta} \cdots \\ &\xrightarrow{\delta} D_1 G^* \otimes \wedge^{g+1} F \xrightarrow{\delta} D_0 G^* \otimes \wedge^g F = \wedge^g F \xrightarrow{\varepsilon} \wedge^0 F,\end{aligned}$$

which according to the preceding conventions we shall write as

$$\begin{aligned}\mathcal{C}^0 : 0 \longrightarrow D_{f-g} \otimes \wedge^f &\xrightarrow{\delta} D_{f-g-1} \otimes \wedge^{f-1} \xrightarrow{\delta} \cdots \\ &\xrightarrow{\delta} G^* \otimes \wedge^{g+1} \xrightarrow{\delta} \wedge^g \xrightarrow{\varepsilon} R.\end{aligned}$$

Here the ε map on the right may, as indicated, be identified with $\wedge^g \varphi : \wedge^g F \to \wedge^g G = R$, the last identification being made by sending γ to 1. Its image is the ideal $I_g(\varphi)$, and the Eagon-Northcott complex should be thought of as an approximation to a resolution of $R/I_g(\varphi)$; we shall see

that it is actually a resolution iff $I_g(\varphi)$ contains a regular sequence of length $f-g+1$.

In the Eagon-Northcott complex the splice map ε comes at the end, so it is not so clear that two complexes are being spliced. In fact the complex on the left is $K(\varphi')^*_{f-g}$, while the one on the right is $K(\varphi')_0$, the complex $0 \to R \to 0$. In the next example, the Buchsbaum-Rim complex

$$\begin{aligned}\mathcal{C}^1 : 0 \longrightarrow D_{f-g-1} \otimes \textstyle\bigwedge^f \xrightarrow{\delta} D_{f-g-2} \otimes \textstyle\bigwedge^{f-1} \xrightarrow{\delta} \cdots \\ \xrightarrow{\delta} \textstyle\bigwedge^{g+1} \xrightarrow{\varepsilon} F \xrightarrow{\partial} G,\end{aligned}$$

the splice map splices $K(\varphi')^*_{f-g-1}$ to $K(\varphi')_1$, the complex $0 \to F \to G \to 0$. The Buchsbaum-Rim complex should be thought of as an approximation to a resolution of the cokernel of the map $\varphi : F \to G$. It is actually a resolution iff $I_g(\varphi)$ contains a regular sequence of length $f-g+1$.

All the complexes $\mathcal{C}^i$ are given in Figure A2.6. The maps in the part below the diagonal line are all ∂; in fact, this part is nothing but $K(\varphi)$, decomposed according to degree in S. The maps along the diagonal line are all ε, while the maps above the diagonal line are all δ. The picture is self-dual: The complex $\mathcal{C}^i$ is dual to the complex $\mathcal{C}^{f-g-i}$. We shall mostly be concerned with the complexes outside the gray region, since, as we shall see, these are often resolutions.

Theorem A2.10. *Let*

$$\mathcal{C} : 0 \to F_e \xrightarrow{\varphi_e} F_{e-1} \xrightarrow{\varphi_{e-1}} \cdots \xrightarrow{\varphi_1} F_1 \xrightarrow{\varphi_0} F_0$$

be one of the $\mathcal{C}^i, i \geq -1$, *and let* $r(j) = \sum_{i=j}^{e} (-1)^{i-j} \operatorname{rank} F_i$, *the rank that* φ_j *would have if* $\mathcal{C}$ *were a resolution.*

a. $\mathcal{C}$ *is a complex.*

b. *For each* $e \geq j \geq 1, \varphi_j$ *has rank* $\leq r(j)$, *and the ideal* $I_{r(j)}(\varphi_j)$ *is contained in and has the same radical as the ideal* $I_{s(j)}\varphi$, *where* $s(j) = \min(g, f-j+1)$.

c. $\mathcal{C}$ *is a free resolution whenever*

$$\operatorname{depth} I_m(\varphi) \geq f-m+1 \quad \textit{for all } m \textit{ with } g \geq m \geq f-e+1,$$

In particular, $\mathcal{C}$ *is a free resolution in the "generic" case, where* φ *is represented by a matrix of indeterminates.*

Remark: Because the $\mathcal{C}^i$ with $i < -1$ are dual to $\mathcal{C}^i$ with $i > -1$, we see that all the $\mathcal{C}^i$ are actually complexes. If the map $\varphi : F \to G$ is a split epimorphism, that is if $I_g(\varphi) = R$, then all the $\mathcal{C}^i$ are split-exact for $i \geq -1$, and by duality again, all the $\mathcal{C}^i$ are split-exact (as the upcoming argument would also show directly). This is actually the only case in which a $\mathcal{C}^i$ is exact for $i < -1$; see Exercise A2.16.

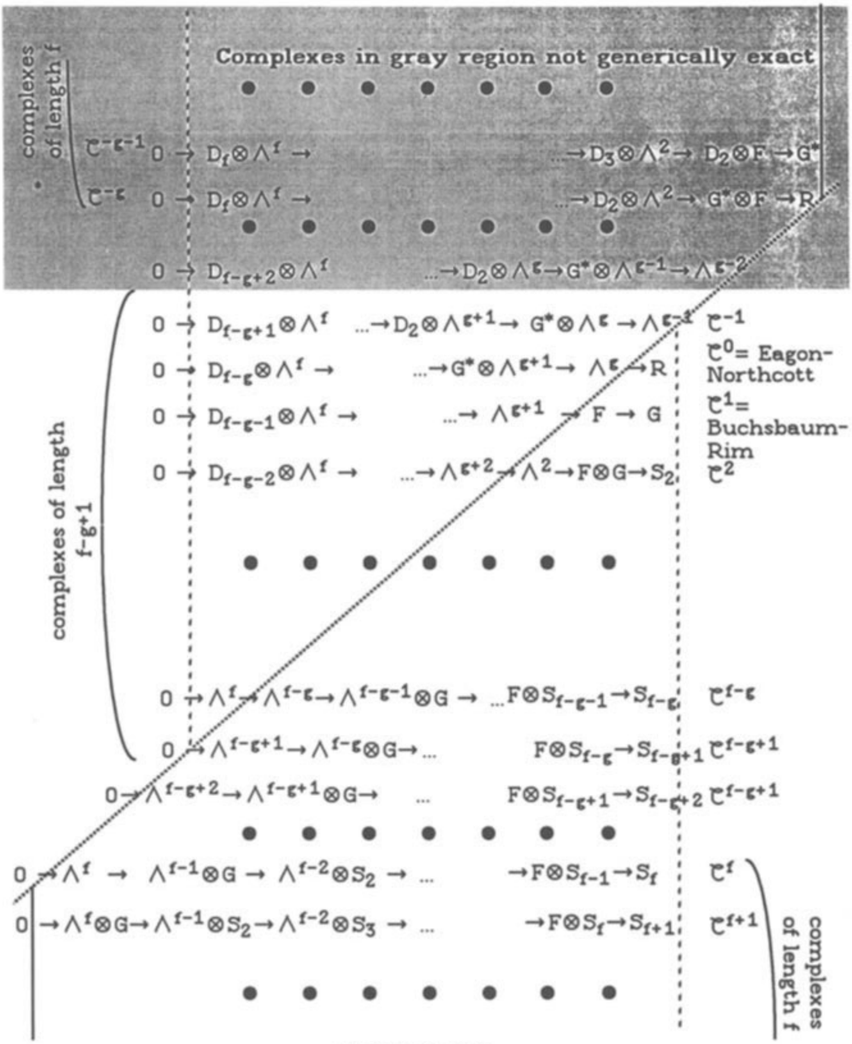


FIGURE A2.6.

An initially surprising feature of the following argument is that modulo several uses of the criterion of exactness (Theorem 20.9), Theorem A2.10 is reduced to results on multilinear algebra that do not involve ring theory at all, but are essentially combinatorial in nature.

Proof. Part c follows from parts a and b by the criterion of exactness (Theorem 20.9).

For parts a and b, since $\mathcal{C}^{-1}$ is dual to $\mathcal{C}^{f-g+1}$, it suffices to prove the theorem for $i \geq 0$. Parts a and b express certain identities on polynomials in the entries of a matrix representation for φ. We have been careful to make

all our constructions compatible with base change (that is, tensoring with an R-algebra), so if the desired identities hold for a given map $\varphi : F \to G$ of free R-modules, and if R' is any R-algebra, then they hold for the map $R' \otimes \varphi : R' \otimes F \to R' \otimes G$ of free R'-modules. Thus it actually suffices to treat the case $R = \mathbf{Z}[\{t_{ij}\}_{1 \le i \le g, 1 \le j \le f}]$, where the t_{ij} are indeterminates, and φ is represented by the $f \times g$ matrix with entries t_{ij}. In this case each $I_m(\varphi)$ is a prime ideal, so that to prove $I_{r(j)}(\varphi_j) \subset I_m(\varphi)$, it suffices to show that they have the same radical. This is what we shall actually prove here (and this suffices for the rest of the assertions); for a proof that $I_m(\varphi)$ is prime in the generic case, see Bruns and Vetter [1988; Remark 2.12 and Theorem 6.3].

a. Since we already know that $\partial^2 = 0 = \delta^2$, it suffices to show that $\varepsilon\delta = 0$ and $\partial\varepsilon = 0$ in the sequence

$$G^* \otimes \wedge^{k+1} F \xrightarrow{\delta} \wedge^k F \xrightarrow{\varepsilon} \wedge^{k-g} F \xrightarrow{\partial} \wedge^{k-g-1} F \otimes G.$$

But it is easy to see that this sequence is self-dual: We already know that $\delta : G^* \otimes \wedge^{k+1} F \to \wedge^k F = \partial^* : G^* \otimes \wedge^{f-k-1} F \to \wedge^{f-k} F$ up to the identifications of $\wedge^{f-n} F \cong \wedge^n F^*$ induced by any choice of an orientation in $\wedge^f F^*$. The dual of ε is similarly isomorphic to an ε, via the same identifications: One may check by direct computation that the diagram

$$\underbrace{\wedge^{f-k+g} F \cong \wedge^{k-g} F^* \xrightarrow{\varepsilon^*} \wedge^k F^* \cong \wedge^{f-k} F}_{\varepsilon}$$

commutes. Thus it suffices to show that $\partial\varepsilon = 0$. Since $R = \mathbf{Z}[t_{ij}]$ is a domain, it is enough to prove that $\partial\varepsilon = 0$ after tensoring with the quotient field of R. We may thus suppose that $F = F' \oplus G$, and $\varphi : F \to G$ is the projection onto the second factor. Consider the diagram in Figure A2.7.

$$\begin{array}{ccccc}
\wedge^k F & \xrightarrow{\varepsilon} & \wedge^{k-g} F & \xrightarrow{\partial} & \wedge^{k-g-1} F \otimes G \\
\wr\Vert & & \Vert\wr & & \\
\sum_{p+q=k} \wedge^p F' \otimes \wedge^q G & \longrightarrow & \sum_{p+q=k} \wedge^p F' \otimes \wedge^q - g G & \longrightarrow & \sum_{p+q=k} \wedge^p F' \otimes \wedge^{q-g-1} G \otimes G \\
\downarrow & & = \quad \uparrow & & \downarrow \\
\wedge^{k-g} F' \otimes \wedge^g G & \longrightarrow & \wedge^{k-g} F' \otimes \wedge^0 G & \longrightarrow & 0
\end{array}$$

FIGURE A2.7.

Here the isomorphisms are those of Proposition A2.2c. The vertical maps at the bottom are the projection to and inclusion of a direct summand. The left-hand horizontal maps are induced by the action

of $\gamma \in \wedge^g G^*$, while the right-hand horizontal maps come from the diagonal map on $\wedge G$. The commutativity is easy to check, and proves that in this case the composition $\partial\varepsilon$ is 0. Even better, it proves that if $\varphi : F \to G$ is a surjection, then the sequence

$$\wedge^k F \xrightarrow{\varepsilon} \wedge^{k-g} F \xrightarrow{\partial} \wedge^{k-g-1} F \otimes G$$

is exact.

b. The heart of the argument is the following lemma, which describes what happens if we invert a minor of φ and factor out the larger minors. The results we want follow formally from the criterion of exactness (Theorem 20.9) once we know this. □

Lemma A2.11. *Let $\varphi, \mathcal{C} = \mathcal{C}^i$, and φ_j be as in Theorem A2.10. If $I_s(\varphi) = R$ but $I_{s+1}(\varphi) = 0$, then*

a. If $s = g$, then $\varphi_1 : F_1 \to F_0$ is a surjection and $\mathcal{C}$ is exact at F_t for all t.

b. If $s < g$, then $\mathcal{C}$ is exact at F_t iff $t > f - s$.

Proof. It is enough to prove the lemma after localizing at a maximal ideal. After such a localization, we may assume that some $s \times s$ minor of φ is a unit and all larger minors are 0. We may thus write $F = F' \oplus G', G = G' \oplus G''$, with rank $F' = f - s$, rank $G' = s$, and rank $G'' = g - s$, and assume that φ is the projection to G'.

By duality it suffices to treat the cases where $t \leq i$: Indeed, if $i \leq f-g+1$ then $\mathcal{C}^i$ is the dual of $\mathcal{C}^{f-g-i}$, and $f-g-i \geq -1$, while if $t > i$ then $F_t = 0$, and there is nothing to prove.

First consider the special case $t = i \leq f - g$, so that $\varphi_{t+1} = \varepsilon$. If $s = g$, then $\varphi : F \to G$ is a surjection, and we established in the proof of part a of Theorem A2.10 that $\mathcal{C}$ is exact at F_t. If, on the other hand, $s < g$, then $\varepsilon = 0$, and $\mathcal{C}$ is not exact at F_t. These assertions correspond to a and b of the lemma in this case.

In all other cases with $t \leq i$, both φ_{t+1} and φ_t are maps in the strand $K(\varphi')_i$ of the Koszul complex, and we may prove the lemma by analyzing the homology of this Koszul complex.

With respect to a suitable basis $\{x_i\}$ of $G, K(\varphi')$ is the Koszul complex of the sequence of elements $(x_1, \ldots, x_s, 0, \ldots, 0)$. By Proposition 17.9 this is the tensor product of the Koszul complex of $(x_1, \ldots, x_s)$ and the Koszul complex of $(0, \ldots, 0)$. It follows at once that the homology of $K(\varphi')$ may be identified with $\wedge F'(-1) \otimes S(G'')$.

If $j \leq f - g + 1$ then $s = g$ so $G'' = 0$ and the homology of the Koszul complex is $\wedge F'(-1)$. In this case the degree-i strand is exact at F_t for all $t < i$, as desired. In the contrary case, when $j > f - g + 1$, the homology is nonzero at the tth step $F_t = \wedge^t F \otimes S_{i-t} G$ of the degree-i strand iff

$$t \leq \operatorname{rank} F' = f - s = \max(f-g, j-1) = j-1$$

and

$$i \geq t,$$

which proves the lemma. □

We now return to the proof of Theorem A2.10. As remarked earlier, it suffices to treat the case where $R = \mathbf{Z}[\{t_{ij}\}]$, and φ is the map associated to the matrix (t_{ij}). Let D be a $g \times g$ minor of φ. Applying Lemma A2.11 we see that the complex $R[D^{-1}] \otimes_R \mathcal{C}$ is split-exact. Thus the rank of $R[D^{-1}] \otimes_R \varphi_j$ is $r(j)$. Since $\mathbf{Z}[\{t_{ij}\}]$ is a domain, the rank of φ_j is also $r(j)$, proving the first statement of part b.

Next we shall show that $\operatorname{rad}(I_{r(j)}(\varphi_j)) \supset \operatorname{rad}(I_{s(j)}(\varphi))$. Suppose that P is a prime of R containing $I_{r(j)}(\varphi_j)$; we must show that $I_{s(j)}\varphi \subset P$. Let Q be the quotient field of R/P. Note that $Q \otimes_R \mathcal{C}$ is not exact at F_j. It follows from Lemma A2.11 that $I_{s(j)}\varphi \subset P$ as desired.

Since $I_{s(j)}\varphi$ has $\operatorname{codim}(f - s(j) + 1)(g - s(j) + 1) \geq j$, the criterion of exactness (Theorem 20.9) shows that $\mathcal{C}$ is a resolution. From Corollary 20.12 it follows that

$$\operatorname{Radical}(I_{r(j)}(\varphi_j)) \subset \operatorname{Radical}(I_{r(j-1)}(\varphi_{j-1})).$$

To finish the proof we must show that $\operatorname{rad}(I_{r(j)}(\varphi_j)) \subset \operatorname{rad}(I_{s(j)}(\varphi))$. Suppose that P is a minimal prime of $I_{s(j)}(\varphi)$; we must show that $P \supset I_{r(j)}(\varphi_j)$. Note that P does not contain $I_{s(j)-1}(\varphi)$, since this ideal has greater codimension than $I_{s(j)}(\varphi)$ by Exercise 10.10. Let Q be the quotient field of R/P. The map $Q \otimes_R \varphi$ satisfies the condition of Lemma A2.11 with $s = s(j) - 1$, so $Q \otimes_R \mathcal{C}$ is exact at F_t iff $t > f - s(j) + 1 = \max(f - g + 1, j)$. In particular, $Q \otimes_R \mathcal{C}$ is not exact at F_j. But if P did not contain $I_{r(j)}(\varphi_j)$ then it would not contain any $I_{r(j')}(\varphi_{j'})$ for $j' \geq j$, so by Theorem 20.9 $Q \otimes_R \mathcal{C}$ would be exact at F_j. The contradiction shows that P contains $I_{r(j)}(\varphi_j)$ as required. □

Corollary A2.12. *Let $(R, \mathfrak{m})$ be a local ring, and suppose that $\varphi : F \to G$ is a map of free modules of ranks $f \geq g$ over R with $I_1(\varphi) \subset \mathfrak{m}$. If* $\operatorname{depth} I_g(\varphi) = f - g + 1$, *the greatest possible value, then $\mathcal{C}^i$ is the minimal R-free resolution for $-1 \leq i \leq g$. Thus for example the ideal $I_g(\varphi)$ is minimally generated by the $\binom{f}{g}$ distinct $g \times g$ minors of φ.*

Proof. That the $\mathcal{C}^i$ are resolutions is the content of Theorem A2.10. Minimality comes from the way in which the maps are made from φ; if all the entries of φ are in $\mathfrak{m}$, then the same goes for each of the maps ε, δ, and ∂. The last statement follows because the free module F_1 in the resolution $\mathcal{C}^0$, which maps onto $I_g(\varphi)$, has a free basis mapping onto the distinct minors of φ.

A module N over a local ring $(R, \mathfrak{m})$ of dimension d is called a **maximal Cohen-Macaulay module** if $\mathfrak{m}$ contains a regular sequence on N having length d. If R is regular, such modules are free by the Auslander-Buchsbaum theorem. See Chapter 21 for more information.

Corollary A2.13. *Let $(R, \mathfrak{m})$ be a local Cohen-Macaulay ring, and suppose that $\varphi : F \to G$ is a map of free modules of ranks $f \geq g$ over R. Set $M = \operatorname{coker} \varphi$. If $\operatorname{depth} I_g\varphi = f - g + 1$, the greatest possible value, then $R/I_g(\varphi)$ is a Cohen-Macaulay ring. Further, the complex $\mathcal{C}^{-1}$ is a resolution of $\wedge^{f-g}(\operatorname{coker} \varphi^*)$, and for $1 \leq i \leq f-g+1$ the complex $\mathcal{C}^i$ is a resolution of $S_i(M)$. These modules are maximal Cohen-Macaulay $(R/I_g(\varphi))$-modules.*

Proof. By Theorem A2.10, the complexes $\mathcal{C}^i$ for $-1 \leq i \leq f - g + 1$ are resolutions of length $f - g + 1$. If we let

$$\mathcal{C}^i : 0 \to F_e \xrightarrow{\varphi_e} F_{e-1} \xrightarrow{\varphi_{e-1}} \cdots \xrightarrow{\varphi_1} F_1 \xrightarrow{\varphi_0} F_0$$

be one of these $\mathcal{C}^i$, then by the Auslander-Buchsbaum formula coker φ_0 is an R-module of depth equal to $\operatorname{depth} R - (f - g + 1)$. If $i = 0$, then coker φ_0 is $R/I_g(\varphi)$. By the equidimensionality of Cohen-Macaulay rings, $\dim R/I_g(\varphi) = \dim R - \operatorname{codim} I_g(\varphi) = \operatorname{depth} R - \operatorname{depth} I_g(\varphi) = \operatorname{depth} R - (f - g + 1)$, so $R/I_g(\varphi)$ is a Cohen-Macaulay ring.

If $i \geq 1$, then φ_0 is the natural map $\partial : F \otimes S_{i-1}G \to S_iG$, and we know from the right-exactness of the symmetric algebra (Proposition A2.2d) that the cokernel is $S_i(M)$.

If $i = -1$, then φ_0 is the map $\delta : G^* \otimes \wedge^g F \to \wedge^{g-1} F$. By duality in the exterior algebra we may identify this map with the map $G^* \otimes \wedge^{f-g} F^* \to \wedge^{f-g+1} F^*$ induced by φ^*, whose cokernel is $\wedge^{f-g+1}(\operatorname{coker} \varphi^*)$.

By Proposition 20.7, M is annihilated by $I_g(\varphi)$, so M, and with it all the $S_i(M)$ are $(R/I_g(\varphi))$-modules. The module $\wedge^{f-g+1}(\operatorname{coker} \varphi^*)$ is also annihilated by $I_g(\varphi)$, by Exercise 20.9. Thus all the modules in question are maximal Cohen-Macaulay $(R/I_g(\varphi))$-modules. □

Our final result shows that under good circumstances all the modules we have resolved are isomorphic to ideals of $R/I_g(\varphi)$. In Exercise A3.30 we shall show how to use the preceding complexes to give resolutions of the preimages of these ideals in R. In Exercises A2.21 and A2.22 we shall see that these can be used to study geometrically significant examples, such as collections of points on a rational normal curve, elliptic normal curves, and trigonal canonical curves.

Theorem A2.14. *Suppose that $I_g(\varphi)$ has depth $f - g + 1$. Fix bases for F and G, and regard φ as a matrix. Let φ_1 be the submatrix consisting of the first $g - 1$ columns of φ; let φ_2 be the submatrix consisting of the first $g - 1$ rows of φ; and let φ_0 be the upper left $(g - 1) \times (g - 1)$ submatrix of φ, as in Figure* A2.8. *Suppose that both $I_{g-1}(\varphi_1) + I_g(\varphi)$ and $I_{g-1}(\varphi_2) + I_g(\varphi)$ have* depth $> f - g + 1$.

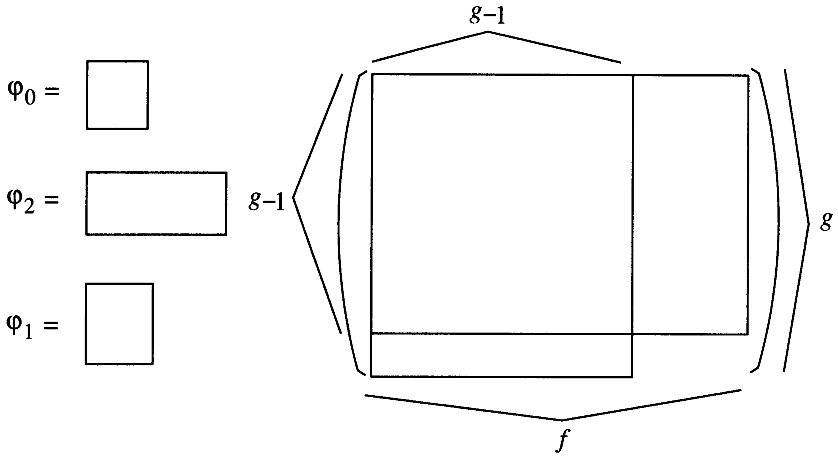


FIGURE A2.8.

a. $M :=$ coker φ *is isomorphic to the ideal* $I = (I_{g-1}(\varphi_1) + I_g(\varphi))/I_g(\varphi)$ *of* $R/I_g(\varphi)$ *generated by the* $(g-1) \times (g-1)$ *minors in the first* $g-1$ *columns of* φ. *For* $j = 1, \ldots, f-g+1$ *we have* $I^j \cong S_j(M)$.

b. *If* $j > f-g+1$ *and* depth $I_{g-t}(\varphi) > f-g+1+t$ *for all* $0 < t \leq j-(f-g+1)$, *then* $I^j \cong S_j(M)$.

c. $N := \wedge^{f-g}(\operatorname{coker} \varphi^*)$ *is isomorphic to the ideal* $J = I_{g-1}(\varphi_2)/I_g(\varphi)$ *generated in* $R/I_g(\varphi)$ *by the* $(g-1) \times (g-1)$ *minors in the first* $g-1$ *rows of* φ.

It can be shown that if depth $I_g(\varphi) = f-g+1$, depth $I_{g-1}(\varphi) \geq f-g+2$, and the ground field is infinite, then after a general row and column operation the conditions of Corollary A2.14 will be satisfied.

Proof. We begin by defining maps $\alpha : M \to R/I_g(\varphi)$ and $\beta : N \to R/I_g(\varphi)$ whose images are the given ideals; these maps are defined without hypotheses on the depths of the various ideals of minors, and they involve only simple determinantal identities, here expressed with multilinear algebra. The depth conditions of the hypothesis will serve to ensure that these maps are monomorphisms.

To describe a map from a module with free presentation

$$X = \operatorname{coker} \psi : H_1 \to H_0$$

to the ring $R/I_g(\varphi)$, it suffices to give a commutative diagram as in Figure A2.9, where we have identified $\wedge^g G$ with R. In the case of the module $X = M$, we have presentation $M = \operatorname{coker} \varphi : F \to G$. Let a be the exterior product of the first $g-1$ basis vectors of F, and let $b = \wedge^{g-1}\varphi(a)$. It is easy to see that the diagram in Figure A2.10 commutes, where the two vertical maps are given by wedging with a and with b respectively. The right-hand

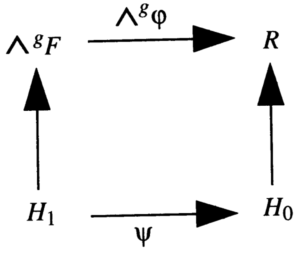


FIGURE A2.9.

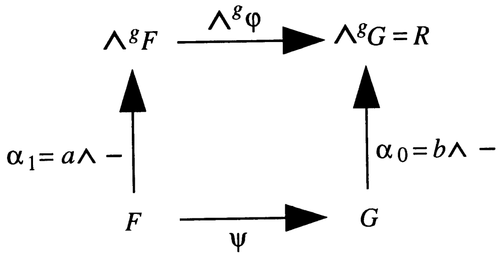


FIGURE A2.10.

vertical map, α_0, sends the ith basis vector $e_i \in G$ to $b \wedge e_i$, which is the $(g-1)\times(g-1)$ minor of φ involving the first $g-1$ columns (corresponding to our choice of a) and all but the ith row. Thus the image of α_0 is the ideal $I_{g-1}(\varphi_1)$. The map α_0 induces a map $\alpha : M \to I$. We deduce maps $S_j(\alpha) : S_j(M) \twoheadrightarrow I^j$ for every j by multiplication.

In the case $X = N$, we have a presentation

$$N = \wedge^{f-g+1}(\operatorname{coker}\varphi^*) = \operatorname{coker}(G^* \otimes \wedge^{f-g}F^* \to \wedge^{f-g+1}F^*).$$

Let $a_1, \ldots, a_g$ be the chosen basis of G, and let $b_1, \ldots, b_g$ be the dual basis of G^*. To give the necessary commutative diagram, we first identify the upper horizontal map $\wedge^g F \to R$ with the map $\wedge^{f-g}F^* \to \wedge^f F^* = R$ that is exterior multiplication by $\wedge^g\varphi^*(b_1 \wedge \cdots \wedge b_g)$. To define the vertical maps in the diagram in Figure A2.11 we set $b = b_1 \wedge \cdots \wedge b_{g-1}$. We define

$$\begin{aligned}\beta_1(c \otimes d) &= \langle c, a_g\rangle d,\\ \beta_0(e) &= \wedge^{g-1}\varphi^*(b) \wedge e,\end{aligned}$$

where $\langle c, a_g\rangle \in R$ is the value of the functional a_g on c. Since the lower horizontal map takes $c \otimes d$ to $\varphi^*(c) \wedge d$, its composite with β takes $c \otimes d$ to $\wedge^g\varphi^*(b \wedge c) \wedge d$. On the other hand, the upper horizontal map takes d to $\wedge^g\varphi^*(b \wedge b_g) \wedge d$. Thus its composite with β_1 takes $c \otimes d$ to $\langle c, a_g\rangle(\wedge^g\varphi^*(b \wedge b_g)) \wedge d$. Since $\langle c, a_g\rangle(b \wedge b_g) = b \wedge c$, the diagram commutes. The image of β_0 is generated by the $(g-1)\times(g-1)$ minors of φ_2. Thus β_0 induces a map $\beta : N \twoheadrightarrow J$.

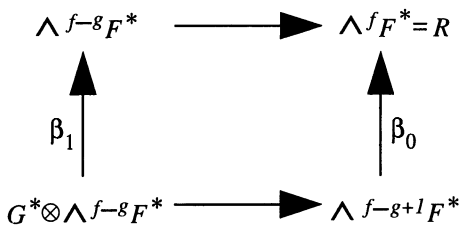


FIGURE A2.11.

We must identify the conditions under which the maps $S_j(\alpha)$ and β are monomorphisms, and we begin with the behavior after certain localizations. Suppose that R is Cohen-Macaulay, that $I_g(\varphi)$ has codimension $f-g+1$, and that both $I_{g-1}(\varphi_1)+I_g(\varphi)$ and $I_{g-1}(\varphi_2)+I_g(\varphi)$ have depth $> f-g+1$, as in the hypothesis of the corollary. $R/I_g(\varphi)$ is Cohen-Macaulay by Corollary A2.13, so $I_g(\varphi)$ is unmixed, and every minimal prime P of $I_g(\varphi)$ has codimension $f-g+1$. Thus in the ring R_P we have $(I_{g-1}(\varphi_1)+I_g(\varphi))R_P = (I_{g-1}(\varphi_2)+I_g(\varphi))R_P = R_P$. We claim that the localizations

$$S_j(\alpha)_P : M_P \to (R/I_g(\varphi))_P,$$
$$\beta_P : N_P \to (R/I_g(\varphi))_P,$$

which we already know are epimorphisms, are in fact isomorphisms. By Proposition 20.7 and Exercise 20.9 both M_P and N_P are annihilated by $I_g(\varphi)_P$, so it suffices to show that each of the modules M_P and N_P can be generated by one element. Proposition 20.6 proves this directly for M_P. The same result shows that $(\operatorname{coker}\varphi^*)_P$ is generated by $f-g+1$ elements, so $N_P = \wedge^{f-g+1}(\operatorname{coker}\varphi^*)_P$ is also generated by one element.

Suppose that L is an $(R/I_g(\varphi))$-module, and that $\gamma : L \to R/I_g(\varphi)$ is any map that is a monomorphism locally at each minimal prime P of $I_g(\varphi)$. The map γ is a monomorphism iff γ is a monomorphism locally at each prime ideal P associated to L by Corollary 3.5. Thus to show that γ is a monomorphism, it is enough to show that the associated primes of L are among the minimal primes of $I_g(\varphi)$.

To complete the proof of parts a and c, let L be either N or one of the modules $S_j(M)$ for $j=1,\dots,f-g+1$. We must show that if $P \in \operatorname{Ass}(L)$ then P is minimal over $I_g(\varphi)$. Since L is resolved by one of the complexes $\mathcal{C}^j$ for $-1 \le j \le f-g+1$, it has projective dimension $\le f-g+1$, so $\operatorname{depth} L \ge \operatorname{depth} R - (f-g+1)$ by the Auslander-Buchsbaum formula. It follows from Corollary 18.5 that $\operatorname{codim} P \le f-g+1$. Since the annihilator of L contains $I_g(\varphi)$, we see that P contains $I_g(\varphi)$. Since $\operatorname{codim} I_g(\varphi) = f-g+1$ by hypothesis, P must be a minimal prime of $I_g(\varphi)$.

For part b, suppose that $j > f-g+1$ and $\operatorname{depth} I_{g-t}(\varphi) > f-g+1+t$ for all $0 < t \le j-(f-g+1)$. By Theorem A2.10, $\mathcal{C}^j$ is a free resolution of $S_j(M)$. By Corollary 20.14b, $\operatorname{Ass}(S_j(M))$ consists entirely of minimal primes of $I_g(\varphi)$, and we are done.

In fact the condition in part b of the corollary is in good cases necessary and sufficient; see Exercise A2.15.

The case that arises most commonly in geometry is that in which $g = 2$, and in which φ is a matrix over a polynomial ring whose entries are linear forms and whose 2×2 minors define a rational normal scroll. In this case the ideals I and J in the proposition define divisors on the scrolls, and the divisors defined by J and I^n for $n = 0, 1, \ldots, f - g + 1$ are, up to adding a multiple of the hyperplane section, the divisors that are themselves arithmetically Cohen-Macaulay varieties. See Eisenbud and Harris [1987] and Schreyer [1986] for more information on this situation. The special cases of sets of points on a rational normal curve, and of an elliptic normal curve, are given in the exercises.

A2.6.2 Exercises

Exercise A2.15: With hypotheses as in the body of Corollary A2.14, suppose that $j > f - g + 1$ and $\operatorname{depth} I_{g-t}(\varphi) \geq f - g + 1 + t$ for all $0 < t \leq j - (f - g + 1)$. Show that $S_j(M)$ is isomorphic to I^j only if $\operatorname{depth} I_{g-t}(\varphi) > f - g + 1 + t$ for all $0 < t \leq j - (f - g + 1)$.

Exercise A2.16:* To convince oneself that at least some of the $\mathcal{C}^i$ in the gray region of Figure A2.6 are not generically exact, note that were $\mathcal{C}^{-g}$ exact, it would be the resolution of the ideal generated by the 1×1 minors of φ, that is, by the entries of a matrix representing φ. But generically the fg entries form a regular sequence, so the ideal should have a resolution of length fg, not length f like the complex $\mathcal{C}^{-g}$. Prove that $\mathcal{C}^i$ for $i < -1$ is *never* a resolution unless the ideal of $g \times g$ minors of φ contains a unit (in which case it is split-exact).

Exercise A2.17 (Matrices of linear forms):

a. Show that the following three kinds of objects are equivalent:

   i. $p \times q$ matrix of linear forms whose entries span a space of linear forms of dimension r in a polynomial ring S over a field k

   ii. an r-dimensional space of $p \times q$ matrices with entries in k

   iii. a pairing $\mu : k^p \otimes k^q \to L$ of k-vector spaces, where L has dimension r and μ is surjective.

b. What matrix of linear forms corresponds to the linear space of all $p \times p$ symmetric matrices? To the linear space of all $p \times p$ matrices of trace 0?

c. Show that the $p \times q$ matrix in Figure A2.12, which is sometimes called the **catalecticant** or **persymmetric** matrix, corresponds to the multiplication pairing $k[s,t]_{p-1} \otimes k[s,t]_{q-1} \to k[s,t]_{p+q-2}$.

$$\begin{pmatrix} x_1 & x_2 & x_3 & \dots & & \\ x_2 & x_3 & x_4 & \dots & & \\ \dots & \dots & \dots & \dots & \dots & \dots \\ & & & \dots & x_{r-2} & x_{r-1} \\ & & & \dots & x_{r-1} & x_r \end{pmatrix} \qquad r = p + q - 1$$

FIGURE A2.12.

d. Find a description related to the one in part c for the pairing that corresponds to the matrix in Figure A2.13.

$$\begin{pmatrix} x_1 & x_2 & \dots & x_r & 0 & \dots & & 0 \\ 0 & x_1 & \dots & x_{r-1} & x_r & 0 & \dots & 0 \\ & & \dots & \dots & \dots & \dots & & \\ 0 & \dots & 0 & 0 & x_1 & x_2 & \dots & x_r & 0 \\ 0 & \dots & & 0 & 0 & x_1 & x_2 & \dots & x_r \end{pmatrix}$$

FIGURE A2.13.

Prove that the ideal generated by the $p \times p$ minors of this matrix is the pth power of the maximal ideal in $k[x_1, \dots, x_r]$; note that this is of codim $r = p + q - 1$, the greatest possible value.

Exercise A2.18 (1-generic matrices): (References: Eisenbud [1988]; Harris [1992].) Let f be a matrix of linear forms over a polynomial ring $S = k[x_1, \dots, x_r]$, with k a field. By a "generalized row" of f we mean a k-linear combination of the rows of f with not all coefficients 0; similarly a generalized column is a nonzero k-linear combination of the columns of f. A generalized entry of f is simply linear combination with nonzero coefficients of the entries of some generalized row. Equivalently, if f represents a map $\varphi : S^p(-1) \to S^q$ of graded free S-modules, then a generalized entry corresponds to a composite map

$$S(-1) \xrightarrow{\alpha} S^p(-1) \xrightarrow{\varphi} S^q \xrightarrow{\beta} S,$$

where α and β are both nonzero. We say that f (or φ) is **1-generic** if every generalized entry is nonzero.

a. Prove that the generic $p \times q$ matrix, defined over the polynomial ring $k[\{x_{ij}\}_{i=1\cdots p, j=1\cdots q}]$ is 1-generic.

b. Prove that the generic symmetric $p \times p$ matrix is 1-generic.

c. Under the correspondence between matrices of linear forms and pairings of vector spaces introduced in Exercise A2.17, show that the pairing $\mu : V \otimes W \to L$ is 1-generic iff for all elements $v \in V$ and

$w \in W$, the element $\mu(v \otimes w)$ is nonzero; that is the kernel of μ does not meet the set of "pure" tensor products $v \otimes w$. We shall call such pairings 1-generic.

d. Use the idea in part c to show that if f is a 1-generic $p \times q$ matrix of linear forms in r variables, then $r \geq p + q - 1$.

e. Prove that the $p \times q$ "catalecticant matrix" in $p + q - 1$ variables is 1-generic.

Parts f and g are for those with some background in algebraic geometry.

f. Generalize the preceding example by showing that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are line bundles on a (reduced irreducible) variety X, and if V and W are vector spaces of sections of $\mathcal{L}$ and $\mathcal{M}$, respectively, then the multiplication pairing

$$\mu : V \otimes W \to H^0(X, \mathcal{L} \otimes \mathcal{M})$$

is 1-generic.

g. Let L be the image of μ. Show that the closure of the image of the rational map $X \to \mathbf{P}^r$ defined by L is contained in the rank-1 locus of the matrix of linear forms corresponding to μ. In the case of the catalecticant map in part c, conclude that the ideal of 2×2 minors of the catalecticant matrix vanishes on the rational normal curve of degree $p + q - 2$ in $\mathbf{P}^{p+q-2}$. In this case the minors generate the ideal of the curve (see Gruson and Peskine [1982]).

Exercise A2.19 (Determinantal varieties associated to matrices of linear forms): Let f be a $p \times q$ matrix of linear forms over $S = k[x_0, \ldots, x_r]$, with $p \leq q$, and consider the set V_m in $\mathbf{P}^r$ defined by the ideal of $(m+1) \times (m+1)$ minors of f—that is, the "rank-m locus" of f.

a. For each generalized row ρ (respectively, column) of f, let L_ρ be the linear subspace of $\mathbf{P}^r$ defined by the vanishing of the linear forms in ρ. Show that V_{p-1} is the union of the L_ρ.

b. If f is 1-generic, use part a to show that the codimension of V_{p-1} is $q - p + 1$, the maximum possible value. Show that in fact if $L \subset \mathbf{P}^r$ is *any* linear space of codim $c < p$, then $L \cap V_{p-1}$ has codim $q - p + 1$ in L.

A slightly harder argument shows that the homogeneous ideal of V_{p-1} is actually generated by the $p \times p$ minors of f, and that the same is true even after cutting with an arbitrary linear space of codim $< p - 1$ (see Eisenbud [1988]). Parts c and d are two special cases, handled by a different technique.

c. Let
$$\varphi = \begin{pmatrix} x_0 & x_1 & \dots & x_{r-1} \\ x_1 & x_2 & \dots & x_r \end{pmatrix}.$$
Show that the ideal I generated by the 2×2 minors of φ is the homogeneous ideal of the rational normal curve parametrized by $(s,t) \mapsto (s^r, s^{r-1}t, \dots, t^r)$, and is thus prime, as follows:

1) Prove that on the open set $x_0 = 1$ the 2×2 minors include the equations $x_i = x_1^i$ that generate the ideal of the curve parametrized by $t \mapsto (1, t, t^2, \dots, t^r)$. Treat the open set $x_r = 1$ similarly. Show that V_{p-1} does not meet the closed set where $x_0 = x_r = 0$. Conclude that the ideal generated by the 2×2 minors differs from the ideal of the rational normal curve at most in a component primary to $(x_0, \dots, x_r)$.

2) Use the Eagon-Northcott complex to show that the projective dimension of S/I is $r-1$, the codimension of I. Conclude from the Auslander-Buchsbaum formula that S/I is Cohen-Macaulay. Now use Macaulay's unmixedness theorem to conclude that I is the ideal of V_{p-1}.

d. Let
$$\varphi = \begin{pmatrix} x_0 & x_1 & \dots & x_{a-1} & y_0 & y_1 & \dots & y_{b-1} \\ x_1 & x_2 & \dots & x_a & y_1 & y_2 & \dots & y_b \end{pmatrix}.$$
Show that the ideal I generated by the 2×2 minors of φ is the homogeneous ideal of the variety, called a rational normal scroll, that is the closure of the set of points of the form $(us^a, us^{a-1}t, \dots, ut^a, vs^b, vs^{b-1}t, \dots, vt^b)$, and is thus prime, by following the same ideas as in part c.

e. (For those who know more algebraic geometry) If f is a 1-generic matrix, representing a map $\varphi : S^p(-1) \to S^q$, consider the variety
$$\tilde{V}_{p-1} = \{(q, \rho) \in \mathbf{P}^r \times \mathbf{P}^p | \text{ if } \rho = (\rho_1, \dots, \rho_p), \text{ then the matrix } \varphi\rho \text{ has entries that vanish at } q\}.$$
Show that $\tilde{V}_{p-1}$ projects to $V_{p-1} \subset \mathbf{P}^r$ and is a resolution of singularities of V_{p-1}. (This resolution is suggested in Room [1938] and may be older; the idea was revived and recast by George Kempf in his thesis.)

Exercise A2.20: Prove a graded version of Theorem A2.14. Assume that R is a polynomial ring and that the matrix φ is a matrix of linear forms. Assuming that the various depth conditions of Theorem A2.14 are satisfied, prove that $I^j \cong S_j(M(-g+1)) = (S_jM)(-j(g-1))$ and $J \cong N(-g+1)$ as graded modules.

Exercise A2.21 (Divisors on rational normal curves): Let $S = k[x_0, \ldots, x_r]$ and let $\varphi : S^r(-1) \to S^2$ be the matrix of linear forms

$$\varphi = \begin{pmatrix} x_0 & x_1 & \ldots & x_{r-1} \\ x_1 & x_2 & \ldots & x_r \end{pmatrix}.$$

Let P be the ideal $I_2(\varphi)$ generated by the 2×2 minors of φ. We have seen in Exercise A2.19 that P is a prime ideal. Set $R = S/P$, and let $I, J \subset R$ be the ideals generated by a column and a row of φ, respectively, say $I = (x_0, x_1)R$ and $J = (x_0, \ldots, x_{r-1})R$.

a. Show that $I \cong (x_t, x_{t+1})R$ for any $0 \le t \le r-1$, and that $J \cong (x_1, \ldots, x_r)R$. (The same would be true for any generalized row and column in the sense of Exercise A2.18.)

b. Show that the hypotheses of Theorem A2.14 are satisfied for φ, so that Exercise A2.20 applies.

c. Show that $\omega_R \cong S_{r-2}(M)(-1) \cong I^{r-2}(r-3)$.

d. (For those who know more algebraic geometry) P is the ideal of the rational normal curve C of degree r in $\mathbf{P}^r$. Show that I is the ideal of a subscheme of length $r-1$ on C, while J is the ideal of one point on C. If D is a divisor of degree $n = dr - j$ on C, show that the ideal of D (in the homogeneous coordinate ring of C) is isomorphic to $S_j(M)(-d)$.

e. Show that the minimality criterion of Exercise A3.30 is satisfied, and with Theorem A2.14 gives a minimal free resolution for the ideal of any set of points on a rational normal curve.

f. Use part d to show that $J^* = I(-1)$ and $I^* = J(-1)$.

Exercise A2.22 (Divisors on a scroll): (For those who know something about linear series on curves and of rational ruled surfaces, say at the level of Hartshorne [1977] Chapters 4 and 5). In this exercise we consider two more applications of Theorem A2.14 in which $f = 2$ (this is almost the only case for which the minors of φ can cut out a nonsingular variety). Using Exercise A3.30, this gives minimal free resolutions of the ideals of elliptic normal curves and of trigonal canonical curves.

Divide the coordinates on $\mathbf{P}^r$ into two groups $x_0, \ldots, x_a$ and $y_0, \ldots, y_b$. We have seen in Exercise A2.19 that the ideal P of 2×2 minors of the matrix

$$\varphi = \begin{pmatrix} x_0 & x_1 & \ldots & x_{a-1} & y_0 & y_1 & \ldots & y_{b-1} \\ x_1 & x_2 & \ldots & x_a & y_1 & y_2 & \ldots & y_b \end{pmatrix}$$

is prime. Let $S = S(a,b) \subset \mathbf{P}^r$ be the corresponding variety, the **rational normal scroll of type** a, b.

a. Show that Exercise A2.20 may be applied to the matrix φ.

b. **(Elliptic normal curves)** Let C be an elliptic curve over an algebraically closed field, and D a divisor on C of degree $r+1$. By the Riemann-Roch theorem we have $h^0(\mathcal{O}_C(D)) = r+1$, and there is a corresponding map $\alpha : C \to \mathbf{P}^r$. If $r \geq 3$, this map is an embedding (see Hartshorne, [1977] Chapter IV). The image of C in $\mathbf{P}^r$ is called an **elliptic normal curve**.

Now let $r \geq 3$, and let $E \subset C$ be any divisor of degree 2. Again by Riemann-Roch, $h^0(\mathcal{O}(\mathcal{E})) = 2$ and $h^0(\mathcal{O}(\mathcal{D} - \mathcal{E})) = r-1$. Using Exercise A2.18 we see that the elliptic normal curve $\alpha(C)$ lies in the rank-one locus of a certain $2 \times r-1$ matrix of linear forms ψ.

We next put ψ into normal form. Show that without loss of generality we may take the divisor E to have the form $2p$, and the divisor D to have the form $(r-2)p+q$, for some points $p, q \in C$. We may represent the space $H^0(\mathcal{O}_C(E))$ as the space of rational functions on C with poles of order at most 2 at p. In this sense we may choose $1, x$ as a basis, where x is a rational function with a pole of order 2. Similarly, we may take a basis for $H^0(\mathcal{O}_C(D-E))$ to be $1, x, x^2, \ldots, x^{a-1}, y, yx, \ldots, yx^{b-1}$, where y is a rational function with a double pole at p and an additional simple pole at q (or a triple pole at p if $q = p$), and a and b are the greatest integers such that the order of the pole of x^{a-1} is $2(a-1) \leq r-1$, and similarly the sum of the orders of the poles of yx^{b-1} is $2b+1 \leq r-1$. Show that with a suitable choice of coordinates on $\mathbf{P}^r$, the matrix ψ is equal to the matrix φ above, so that the elliptic normal curve $\alpha(C)$ is a divisor on the rational normal scroll S of type a, b.

The divisor class group of the scroll is $\mathbf{Z} \oplus \mathbf{Z}$, generated by the class of a hyperplane H and the class of a ruling F. Show that the elliptic normal curve $\alpha(C)$ has class $2H - (r-3)F$, so that the ideal of $\alpha(C)$ in the scroll is $\mathcal{O}_S((r-3)F - 2H)$. Show that the elements of any column of the matrix φ vanish on a ruling F of the scroll (that is, a plane of dimension $r-2$ contained in the scroll).

With notation of Theorem A2.14, show that the ideal of $\alpha(C)$ in the scroll is isomorphic to $S_{r-3}(M)(-2)$. Show that the minimality criterion of Exercise A3.30 is satisfied, so that the mapping cone described there is a minimal free resolution of the ideal of the elliptic normal curve in $\mathbf{P}^r$.

(The most familiar case: If $r = 3$, then $\alpha(C)$ is a complete intersection of two quadrics in $\mathbf{P}^3$. See the analysis following Corollary 18.14.)

c. **(Trigonal canonical curves)** Now let C be a curve of genus $r \geq 4$ that is **trigonal** in the sense that there is a base-point free divisor E of degree 3 on C such that $h^0(\mathcal{O}_C(E)) = 2$. Let K be a canonical

divisor on C. The complete canonical system $H^0(\mathcal{O}_C(K))$ defines an embedding $\alpha : C \to \mathbf{P}^r$. Applying the ideas of part a to the divisors E and $K - E$, show that C lies on a rational normal scroll S of type a,b where a is the greatest integer such that $K - (a+1)E$ is effective. Show that the divisor class of $\alpha(C)$ in S is $3H - (r-3)E$. Show as in part a that Theorem A2.14 (in the form of Exercise A2.20) and Exercise A3.30 lead to the description of the minimal free resolution of the ideal of $\alpha(C)$ in $\mathbf{P}^r$.

(The most familiar case: If $r = 3$, then $\alpha(C)$ is a complete intersection of a quadric and a cubic in $\mathbf{P}^3$. It is not hard to show that any nonsingular curve that is the complete intersection of a quadric and a cubic in $\mathbf{P}^3$ is a canonical curve (necessarily trigonal because the genus is 4).

Appendix 3
Homological Algebra

A3.1 Introduction

A **complex** of modules over a ring R is a sequence of R-modules and homomorphisms

$$\mathcal{F}: \quad \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots$$

such that $\varphi_i \varphi_{i+1} = 0$ for each i. The **homology** of F at F_i is defined to be

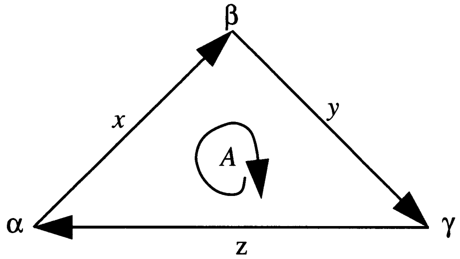
$$H_i\mathcal{F} := \ker \varphi_i / \operatorname{im} \varphi_{i+1}.$$

Homological algebra is, roughly speaking, the study of complexes of modules and their homology.

Some basic terminology: The module F_i is called the **term of degree *i*** of F. For reasons we shall explain the maps φ_i are often called the "boundary operators," or "differentials," of $\mathcal{F}$. The elements of the image of φ_{i+1} are accordingly called boundaries, and the elements of the kernel of φ_i are called cycles. We think of $\mathcal{F}$ as having infinitely many terms, but we shall almost always be concerned only with complexes where $F_i = 0$ either for all $i < 0$ or for all $i > 0$. It is often convenient not to indicate explicitly the terms that are zero. The complex $\mathcal{F}$ is said to be **exact** at F_i if $H_iF = 0$; we say that $\mathcal{F}$ is exact if it is exact at every F_i.

Complexes appear in the work of Cayley fairly explicitly as early as [1858]. They were used by Hilbert in his famous work [1890] to compare a factor ring of a polynomial ring to the polynomial ring itself (just as we

shall use free resolutions to compare an arbitrary module to free modules); the context of his application was explained in Chapter 1. Our current terminology was introduced much later. The name "complex," for example, arose from the simplicial complexes of topology: To an oriented simplicial complex Poincaré [1899] associated a "chain complex," with geometrically defined "boundary operator." The case of a triangle is illustrated in the figure. The formulation in terms of groups and maps came later, apparently suggested by Emmy Noether to several people in the mid-1920s.



$$\begin{aligned} \mathbf{Z}A &\to \mathbf{Z}x \oplus \mathbf{Z}y \oplus \mathbf{Z}z \to \mathbf{Z}\alpha \oplus \mathbf{Z}\beta \oplus \mathbf{Z}\gamma \\ \partial A &= x + y + z, \\ \partial x &= \beta - \alpha,\ \partial y = \gamma - \beta,\ \partial z = \alpha - \gamma \end{aligned}$$

Another part of the prehistory of homological algebra is Poincaré's study of the complex of differential forms on a manifold that we now call the de Rham complex: Poincaré's lemma asserts that the de Rham complex of $\mathbf{R}^n$ is exact. (de Rham's name attaches to the complex because he was the first to prove—in the 1940s—that the cohomology of the de Rham complex is a topological invariant.) The maps in the de Rham complex are derived from differentiation, and it was natural to call them differentials:

$$\mathcal{C}^\infty_{\mathbf{R}} \xrightarrow{\partial} T^*_{\mathbf{R}} \to \wedge^2 T^*_{\mathbf{R}} \to \cdots \qquad \partial(f(x)) = f'(x)\,dx.$$

If M is an R-module, we may consider M as a complex

$$\cdots \to 0 \to M \to 0 \to \cdots$$

with only one nonzero term. Thus homological algebra includes the study of modules. In commutative algebra, homological algebra is usually pursued in order to study modules more closely. Complexes give us a way of comparing an arbitrary module with nicer ones—with free, or projective, or injective modules. Perhaps the most complete expression of this idea is in the construction of the derived category, which we describe briefly in the last section of this appendix.

Complexes arise naturally from the study of systems of linear equations: A system of n_0 linear equations in n_1 unknowns over a ring R corresponds

to an $n_0 \times n_1$ matrix φ over R (the matrix of coefficients of the equations), or alternately as a map of free R-modules, $\varphi : F_1 = R^{n_1} \to R^{n_0} = F_0$. A family of solutions to the (homogeneous) equations

$$\varphi X = 0$$

may be described by a map $F_2 \to F_1$ making the sequence

$$F_2 \to F_1 \to F_0$$

a complex. Solving the equations means giving a "complete" set of solutions; that is, a map as above making the complex exact at F_1.

If R is a field, then of course there is a finite linearly independent set of solutions in terms of which all others can be expressed. For more general rings, this is no longer the case: It may be impossible to choose a generating set for the kernel consisting of linearly independent elements.

An example will make the situation clear. Let $R = k[a,b,c]$ be a polynomial ring in three variables, and consider the linear equation in three unknowns

$$aX_1 + bX_2 + cX_3 = 0$$

corresponding to the map $\varphi : R^3 \to R$ with matrix (a,b,c). By analogy with our experience of linear equations over a field, we should say that the rank of this system is 1, so we should expect $3 - 1 = 2$ independent solutions. However, the three columns of the matrix

$$\begin{pmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{pmatrix}$$

are all solutions (elements of ker φ). It is easy to see that they actually generate ker φ, but that no two elements generate it. Furthermore, these three generators are linearly dependent in the sense that if we multiply the first column by a, the second by b, and the third by c, and add, we get 0. It is not hard to show (see Chapter 20) that every complete set of solutions must be linearly dependent. Thus we have a situation that could not have arisen over a field: a system of linear equations such that any complete set of solutions is linearly dependent.

If we wish to describe the solutions to our original system of equations as linear combinations of the solutions in a complete set of solutions, then we must describe the linear dependencies (otherwise, we won't be able to tell which linear combinations give the trivial solution). If we have n_2 solutions, and we define a new map φ_2 of free R-modules $F_2 = R^{n_2}$ to F_1 by sending the basis elements of F_2 to our solutions, then the dependencies are the elements of the kernel of φ_2. We may regard φ_2 as being a new system of linear equations, and the process of solving begins again. With hindsight we rename φ as φ_1, and continuing the process above we finally arrive at a complex:

$$\cdots \to R^{n_i} \xrightarrow{\varphi_i} \cdots \xrightarrow{\varphi_2} R^{n_1} \xrightarrow{\varphi_1} R^{n_0};$$

in fact, this is an especially interesting sort of complex, called a **free resolution**. In the example above, the resolution actually ends at the next step beyond the one we have given, in the sense that the kernel of φ_2 is itself free (that is, φ_2 has a complete system of solutions with no dependencies). It may be exhibited as in Figure A3.1.

$$0 \longrightarrow R \xrightarrow{\begin{pmatrix} a \\ b \\ c \end{pmatrix}} R^3 \xrightarrow{\begin{pmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{pmatrix}} R^3 \xrightarrow{(abc)} R$$

FIGURE A3.1.

This phenomenon is typical of rings called regular rings; see Chapter 19. The complex given here is called a **Koszul complex** (the name, though universal, is misleading: Such complexes appeared in the works of Cayley and Hilbert before Koszul was born).

We shall now take up these notions systematically, if somewhat sketchily. The proofs that we have omitted are all easy, and we leave them as exercises for the reader.[1] The goal of the first half of this appendix is the theory of derived functors; Ext, Tor, and local cohomology are the most important ones here. One of our less traditional topics in this part is the theory of injective modules over a Noetherian ring. The second half of the appendix is an introduction to spectral sequences.

As everywhere in this book, we shall work with modules over a commutative ring, but the reader should know that nearly everything here can be generalized with just a little effort to modules over an arbitrary ring, or even to objects in a nice Abelian category. Jans [1964], Rotman [1979], and Maclane [1963] are readable sources for more information, roughly in order of increasing difficulty and comprehensiveness. The book of Gelfand and Manin [1989] should soon be available in English.

Part I: Resolutions and Derived Functors

Let R be a commutative ring; the modules in this chapter will all be R-modules.

---

[1]This is not so bad. A famous exercise from Serge Lang's influential textbook *Algebra* [Addison-Wesley, Reading, MA, 1965, p. 105] reads: "Take any book on homological algebra, and prove all the theorems without looking at the proofs given in that book."

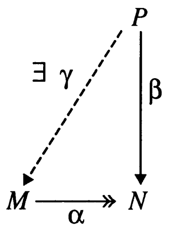
A3.2 Free and Projective Modules

The easiest modules to understand are the free modules: direct sums of copies of the ring. From our point of view free modules are useful because it is easy to define a map from a free module: Namely, suppose F is free on a set of generators p_i (that is, $F \cong \oplus_i R$, and we denote the generator of the ith summand by p_i). To define a map from F to any module M it is enough to tell where to send the generators p_i, and any choice of images for these elements will do. That is,

$$\mathrm{Hom}_{R-\mathrm{modules}}(F, M) = \mathrm{Hom}_{\mathrm{Sets}}(\{p_i\}, M).$$

(In the language of category theory (Appendix 5), the "free module functor is left-adjoint to the forgetful functor"; but we shall not use this formulation here.) This property makes them **projective** in the following sense:

Definition. *A module P is projective if for every epimorphism of modules $\alpha : M \twoheadrightarrow N$ and every map $\beta : P \to N$, there exists a map $\gamma : P \to M$ such that $\beta = \alpha\gamma$, as in the following figure.*



Free modules are projective because if P is free on a set of generators p_i, then we may choose elements q_i of M that map to the elements $\beta(p_i) \in N$, and take γ to be the map sending p_i to q_i.

The definition of projectivity has several useful reformulations:

Proposition A3.1. *Let P be an R-module. The following are equivalent:*

a. *P is projective.*

b. *For every epimorphism of modules $\alpha : M \twoheadrightarrow N$, the induced map $\mathrm{Hom}(P, M) \to \mathrm{Hom}(P, N)$ is an epimorphism.*

c. *For some epimorphism $F \twoheadrightarrow P$, where F is free, the induced map $\mathrm{Hom}(P, F) \to \mathrm{Hom}(P, P)$ is an epimorphism.*

d. *P is a direct summand of a free module.*

e. *Every epimorphism $\alpha : M \twoheadrightarrow P$ "splits": That is, there is a map $\beta : P \to M$ such that $\alpha\beta = 1_P$.*

Proof.

a ⇔ b: This is a restatement of the definition.

b ⇒ c: Obvious.

c ⇒ d: Any map $\varphi \in \mathrm{Hom}(P, F)$ in the preimage of the identity map $1 \in \mathrm{Hom}(P, P)$ is a splitting of the epimorphism $F \twoheadrightarrow P$, so P is a summand of F.

d ⇒ b: This follows because for any modules P and Q we have

$$\mathrm{Hom}(P \oplus Q, -) = \mathrm{Hom}(P, -) \oplus \mathrm{Hom}(Q, -).$$

We have now shown that a, b, c, and d are equivalent.

a ⇒ e: Apply the definition in the case where β is the identity map of P.

e ⇒ d: Obvious. □

As a first example, the reader may check that a finitely generated **Z**-module is projective iff it is torsion-free iff it is free.

Not all projective modules are free; perhaps the simplest example is the ideal $(2, 1+\sqrt{-5}) \subset \mathbf{Z}[\sqrt{-5}]$; see Chapter 11. In general, whether or not a projective module is free is quite a hard question. (We have already discussed the connection of this question to number theory.)

Geometrically, projective modules correspond to algebraic vector bundles: The set of sections of a vector bundle on a variety X is a module over the ring of regular functions on X. This connection is sketched in Corollary A3.3. The relation of algebraic vector bundles to the structure of X is more subtle than in the topological case. For example, topological vector bundles on contractible spaces are easily shown to be trivial, but it was only recently shown (by Quillen and Suslin, in answer to a celebrated problem of Serre; see Lam [1978]) that algebraic vector bundles on $\mathbf{A}^n_k$ are trivial—that is, that projective modules over $k[x_1, \ldots, x_r]$, where k is a field, are free.

Projective modules behave well under localization. Moreover, there is a useful "local criterion" for projectivity, established in Exercises 4.11 and 4.12 and their hints, and summarized as follows:

Theorem A3.2 (Characterization of projectives). *Let M be a finitely presented module over a Noetherian ring R. The following are equivalent:*

a. M is a projective module.

b. M_P is a free module for every maximal ideal (and thus for every prime ideal) P of R.

c. *There is a finite set of elements $x_1, \ldots, x_r \in R$ that generate the unit ideal of R, such that $M[x_i^{-1}]$ is free over $R[x_i^{-1}]$ for each i.*

In particular, every projective module over a local ring is free. Every graded projective module over a positively graded ring R with R_0 a field is a graded free module. □

Corollary A3.3. *Finitely generated projective modules over the affine ring A of a variety X correspond to vector bundles on X: Given a vector bundle E, its sections $\Gamma(E)$ form a finitely generated projective A-module, and any finitely generated projective module arises from a unique vector bundle in this way.*

Proof (*sketch, for those who know about sheaves*). If E is a vector bundle, then there is a covering of X by affine open sets $X_i = \{p \in X | x_i(p) \neq 0\}$ such that $E|_{X_i}$ is trivial. Thus $\Gamma(E|_{X_i}) = \Gamma(E)[x_i^{-1}]$ is free, and $\Gamma(E)$ is projective by Theorem A3.2. Conversely, suppose M is a finitely generated projective module. By Theorem A3.2 we may find elements $x_1, \ldots, x_r$ that generate the unit ideal and such that $M[x_i^{-1}]$ is free (of some rank r_i) for each i. Let E_i be the trivial bundle on X_i of rank r_i. Choose an isomorphism $\alpha_i : \Gamma(E_i) \to M[x_i^{-1}]$. On $X_i \cap X_j$ we may form the composite

$$\Gamma(E_{i|X_j}) \xrightarrow{\alpha_i} M[(x_i x_j)^{-1}] \xrightarrow{\alpha_j^{-1}} \Gamma(E_{j|X_i}),$$

and this determines an isomorphism of bundles $a_{ij} : E_{i|X_i \cap X_j} \to E_{j|X_i \cap X_j}$. Using the maps a_{ij} as gluing maps, we reconstruct a vector bundle E on X. An easy computation using Exercise 2.19 shows that $M = \Gamma(E)$. Further, if M is the module of sections of a vector bundle E' to begin with, then the identification of modules of sections

$$\Gamma(E_i) \to M[x_i^{-1}] = \Gamma(E'_{|X_i})$$

comes from an isomorphism $E_{|X_i} = E_i \to E'_{|X_i}$. Since these isomorphisms are compatible with the gluings, we get $E' \cong E$. □

A3.3 Free and Projective Resolutions

As we have already noted, every module M is an epimorphic image of a free (and thus projective) module—just choose a set of generators $\{g_i\}$ for M, and map a free module on a corresponding set of generators $\{e_i\}$ to M by sending e_i to g_i. This makes it easy to compare any module to free modules: if $\alpha : F_0 \twoheadrightarrow M$ is an epimorphism, then we may say that F_0 differs from M by the module $\ker \alpha$. We may thus express M in terms of free

modules "better" by mapping a free module F_1 onto $\ker\alpha$. Taking φ_1 to be the composite

$$F_1 \twoheadrightarrow \ker\alpha \rightarrowtail F_0,$$

we may say instead that $M = \operatorname{coker}\varphi_1 : F_1 \to F_0$. Unfortunately, there is still a (possibly) nonfree module lurking in this description: the kernel of φ_1. We can repair this defect to some extent by taking a free module F_2 that maps onto $\ker\varphi_1$. Writing $\varphi_2 : F_2 \to F_1$ for the composite

$$F_2 \twoheadrightarrow \ker\varphi_1 \hookrightarrow F_1,$$

we may think of M as given by the sequence of free modules

$$F_2 \to F_1 \to F_0.$$

There is still the problem that $\ker\varphi_2$ might not be free. Repeating the process above indefinitely if necessary, we may at last obtain a sequence of free modules

$$F : \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots \xrightarrow{\varphi_1} F_0$$

with the properties that φ_{i+1} maps F_{i+1} onto the kernel of φ_i for each $i \geq 1$, and that M is the cokernel of φ_1. A sequence of free modules F_i and maps φ_i with these properties is called a **free resolution** of M. If the F_i are merely projective, it is called a **projective resolution**. Note that F is a complex in the sense above (regarding all the F_i with $i < 0$ as 0).

Example. Perhaps the simplest nontrivial, finite free resolutions are the Koszul complexes; see Chapter 17. The simplest nontrivial, infinite resolution might be the following:

Let $S = k[x]$ be a polynomial ring in one variable (k could be any ring, but might as well be taken to be a field). Let $R = S/(x^n)$, and let M be R/Rx^m, with $0 < m < n$, regarded as an R-module. Here is a free resolution of M as an R-module:

$$\cdots \xrightarrow{x^{n-m}} R \xrightarrow{x^m} R \xrightarrow{x^{n-m}} R \xrightarrow{x^m} R,$$

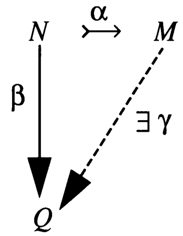
where we have written x^a for the map that is multiplication by x^a. We leave the easy verification to the reader.

A3.4 Injective Modules and Resolutions

The notion of an **injective** module is dual to that of a projective module, but perhaps because injective modules are almost never finitely generated, they are not so familiar.

Definition. *An R-module Q is* ***injective*** *if for every monomorphism of R-modules $\alpha : N \rightarrowtail M$ and every homomorphism of R-modules $\beta : N \to Q$,*

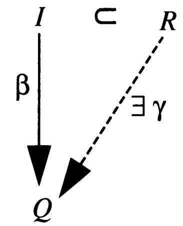
there exists a homomorphism of R-modules $\gamma : M \to Q$ such that $\beta = \gamma\alpha$, as in the figure.



Although the definition of injective modules is precisely dual to that of projective modules, the theory is not dual at all (the category of modules is quite different from its dual category, so this should not be a surprise). The subject is quite beautiful, and we shall explain its beginning.

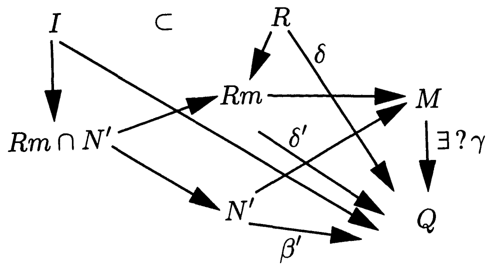
We begin with a result of Reinhold Baer (who defined injective modules in [1940]), showing that in the definition of injectives it is enough to check the case where α is the inclusion of an ideal in the ring.

Lemma A3.4 (Baer). *Let Q be an R-module. If for every ideal $I \subset R$, every homomorphism $\beta : I \to Q$ extends to R as in the diagram, then Q is injective.*



Proof. Suppose M and N are arbitrary R-modules. Let $\beta : N \to Q$ be a homomorphism, and let $\alpha : N \rightarrowtail M$ be a monomorphism. If N' is a submodule of M containing N, then we shall say that $\beta' : N' \to Q$ is an extension of β if β' restricts to β on N. We wish to show that there is an extension of β to M. By Zorn's lemma, there is a submodule N' and an extension β' of β to N' that is maximal in the sense that β' can be extended no further. If $N' = M$, we are done.

Supposing that $N' \neq M$, we shall derive a contradiction. Let $m \in M$ be outside of N', and consider the submodule $N'' = N' + Rm$. Let $I = \{r \in R | rm \in N'\}$. By hypothesis the map $I \to Q$ sending $r \in I$ to $\beta'(rm) \in Q$ extends to a map $\delta : R \to Q$. The map δ induces a map $\delta' : Rm \to Q$ because the kernel of the map $R \to Rm$ is contained in $\ker \delta$, and δ' agrees with β' on $Rm \cap N'$ by definition. We may thus define an extension β'' of β' to N'' by letting β'' be β' on N' and δ' on Rm. This contradicts the maximality of N' and β'. All the necessary maps are shown in the figure. $\square$



Injective modules over $\mathbf{Z}$—that is, injective Abelian groups—are easy to describe:

Proposition A3.5. *An Abelian group Q is injective iff it is divisible in the sense that for every $q \in Q$ and every $o \neq n \in \mathbf{Z}$ there exists $q' \in Q$ such that $nq' = q$.*

Proof. Let Q be injective, and let $q \in Q, o \neq n\ n \in \mathbf{Z}$. Let $\beta : \mathbf{Z} \to Q$ be the map sending 1 to q and let $\alpha : \mathbf{Z} \to \mathbf{Z}$ be multiplication by n. Since Q is injective there is a map $\gamma : \mathbf{Z} \to Q$ with $\beta = \gamma\alpha$. It follows that $n\gamma(1) = q$, so Q is divisible.

Conversely, suppose that Q is divisible. We apply Baer's lemma, A3.4: Let $(n) \subset \mathbf{Z}$ be the inclusion of an ideal. Suppose a map $\beta : (n) \to Q$ takes n to q. Since Q is divisible we may choose $q' \in Q$ with $nq' = q$. The map $\gamma : \mathbf{Z} \to Q$ sending 1 to q' obviously extends β.

From Proposition A3.5 we easily derive a result that is dual to the statement that subgroups of free groups are free.

Corollary A3.6. *If Q is an injective Abelian group, and K is any subgroup, then Q/K is an injective Abelian group.*

Proof. If Q is divisible, then Q/K is divisible too.

We can now show that every Abelian group may be embedded in an injective Abelian group:

Corollary A3.7 (Baer [1940]). *There are "enough" injective Abelian groups, in the sense that for every module M there is a monomorphism $i : M \to Q$ with Q injective.*

Proof. Write $M = F/K$, with F a free module. F is contained in the $\mathbf{Q}$-vector space $F \otimes_{\mathbf{Z}} \mathbf{Q}$, which is obviously divisible. Thus M is contained in the divisible group $(F \otimes_{\mathbf{Z}} \mathbf{Q})/K$. □

Remarkably enough, the corresponding statement for modules over any ring, which is the main goal of our development, is an immediate consequence. (In fact the same argument works still more generally, for example in categories of sheaves of modules over a "ringed space"—a fact that is

exploited in the cohomology theory of sheaves. See, for example, Hartshorne [1977, p. 207].) The key observation is this:

Lemma A3.8. *If R is an S-algebra, and Q' is an injective S-module, then $Q := \mathrm{Hom}_S(R, Q')$ is an injective R-module (the R-module structure comes via the action of R on the first factor of $\mathrm{Hom}_S(R, Q')$).*

For a partial converse, see Exercise A3.7a.

Proof. Let $N \subset M$ be a submodule, and let $\beta : N \to Q$ be a homomorphism; we must show that β extends to M. There is a natural map of S-modules $Q \to Q'$, sending a homomorphism φ to $\varphi(1)$. Let β' be the composite of β and this map $Q \to Q'$, and let γ' be an extension of β' to M, regarded as an S-module. We may define the desired map $\gamma : M \to Q$ of R-modules by sending m to the map φ defined by $\varphi(r) = \gamma'(rm)$.

Corollary A3.9. *For any ring R, the category of R-modules has* **enough injective objects**, *in the sense that for every module M there is a monomorphism $i : M \to Q$ with Q injective.*

Proof (Eckmann and Schöpf [1953]). There is a monomorphism $\alpha : M \to \mathrm{Hom}_{\mathbf{Z}}(R, M)$ sending m to the map φ given by $\varphi(r) = rm$. Temporarily viewing M as an Abelian group, we know that there is a monomorphism of Abelian groups, $\beta : M \to Q'$ of M into an injective Abelian group Q'; applying the functor $\mathrm{Hom}_{\mathbf{Z}}(R, -)$ we get a monomorphism $\beta' : \mathrm{Hom}_{\mathbf{Z}}(R, M) \to \mathrm{Hom}_{\mathbf{Z}}(R, Q')$. By Lemma A3.8 the module $\mathrm{Hom}_{\mathbf{Z}}(R, Q')$ is an injective R-module. Thus $\beta'\alpha$ is a monomorphism of M to an injective module, as desired. □

If M is an R-module, then by Corollary A3.9 we may embed M in an injective module Q_0. We may then embed the cokernel, Q_0/M, in an injective module Q_1. Continuing in this way, we get an **injective resolution**

$$0 \to M \to Q_0 \to Q_1 \to Q_2 \to \cdots$$

of M; that is, an exact sequence of the given form in which all the Q_i are injectives. We shall see how such resolutions are used in the upcoming section on derived functors.

Example. The most familiar injective modules are the divisible Abelian groups. Perhaps the simplest interesting injective resolution is that of $\mathbf{Z}$ as a $\mathbf{Z}$-module:

$$0 \to \mathbf{Z} \to \mathbf{Q} \to \mathbf{Q}/\mathbf{Z} \to 0.$$

In general, injective modules have an interesting and simple structure; see the exercises for more information.

In Chapter 20 it is shown that if R is a local ring then every finitely generated R-module has a unique minimal projective (actually free) resolution. The situation for injective modules is much better: Any module

over any ring has a unique minimal injective resolution! The key idea is that of **injective envelope** (or **injective hull**). First a preliminary definition:

Let R be a ring and let $M \subset E$ be R-modules. We say that M is an **essential submodule** of E, or that E is an **essential extension** of M if every nonzero submodule of E intersects M nontrivially.

Proposition–Definition A3.10. *Let R be a ring.*

a. *Given any R-modules $M \subset F$, there is a maximal submodule E of F containing M such that $M \subset E$ is essential.*

b. *If F is injective, then so is E.*

c. *There is, up to isomorphism, a unique essential extension E of M that is an injective R-module; this E is called the **injective envelope** of M, written $E(M)$.*

Proof.

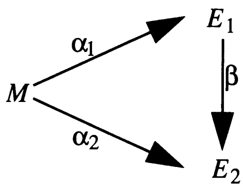
a. If $M \subset E_1 \subset E_2 \subset \cdots \subset F$ with $M \subset E_i$ essential, then any submodule N of $\cup_i E_i$ meets some E_i nontrivially, and thus meets M nontrivially. Thus M is essential in $\cup E_i$. Since M is essential in M, it follows by Zorn's lemma that there exist maximal essential extensions of M contained in F.

b. Suppose now that F is injective, and $M \subset E \subset F$ with $M \subset E$ a maximal essential extension of M by a submodule of F. If E' were an essential extension of E in F, then any nontrivial submodule of E' would meet E, and thus M, nontrivially, so E' would be an essential extension of M and so $E' = E$ by hypothesis. It thus suffices to treat the case where $M = E$. Let N be a submodule of F maximal among those not meeting E; such submodules exist by Zorn's lemma. Since E and N do not meet, we see that $E \oplus N \cong E + N \subset F$. We shall show that $F = E + N$, from which it follows that $E \oplus N \cong F$, so E is injective.

   Consider the composite map $\alpha : E \subset F \twoheadrightarrow F/N$. Because N does not meet E, α is an inclusion. It is essential, for if a submodule N' of F/N failed to meet E, then its preimage in F would be a submodule larger than N and not meeting E, contradicting our hypothesis. Since F is injective, we may find a map $\beta : F/N \to F$ extending α. Since $(\ker \beta) \cap E = \ker \alpha = 0$, and E is essential in F/N, we see that $\ker \beta = 0$. In particular, $\beta(F/N)$ is an essential extension of E. It follows from the maximality of E that $\beta(F/N) = E$, so $F/N = E$, and $E + N = F$ as desired.

c. By **Corollary A3.9** there exist monomorphisms from M to an injective R-module F. From parts a and b we see that a maximal submodule $E \subset F$, such that $M \subset E$ is essential, is injective too.

For uniqueness, suppose that $\alpha_1 : M \to E_1$ and $\alpha_2 : M \to E_2$ are both essential inclusions, with E_1 and E_2 injective, then by the injectivity of E_2 there exists a map $\beta : E_1 \to E_2$ extending α_2 in the sense that the following diagram



commutes. Since $\ker \beta|_M = \ker \alpha_2 = 0$, and $\alpha_1(M)$ is essential in E_1, we see that $\ker \beta = 0$. Thus $\beta(E_1)$ is an injective submodule of E_2. It follows that $E_2 = \beta(E_1) \oplus E_2'$ for some submodule E_2' of E_2. Since $\alpha_2(M)$ is essential in E_2, and $\alpha_2(M) \subset \beta(E_1)$, we must have $E_2' = 0$, and β is the required isomorphism. □

We now say that an injective resolution

$$(*) \qquad 0 \to M \to Q_0 \to Q_1 \to Q_2 \to \cdots$$

is a **minimal injective resolution** if, setting $M_i = \text{coker}(Q_{i-1} \to Q_i)$, we have $Q_{i+1} = E(M_i)$, and the map $Q_i \to Q_{i+1}$ to be the composite of the natural maps

$$Q_i \to M_i \to E(M_i) = Q_{i+1}.$$

As an immediate consequence of Proposition A3.10 we have:

Corollary A3.11. *If R is any ring and M is any R-module, then M has a unique minimal injective resolution.* □

A3.4.1 Exercises

Injective Envelopes

Exercise A3.1: The following principle was used several times in the text: Show that if $N \subset M$ is an essential submodule then any map $M \to E$ of modules that restricts to a monomorphism on N is a monomorphism.

Injective Modules over Noetherian Rings

Exercise A3.2 a. (Bass' Characterization of Noetherian rings):* Show that arbitrary direct products of injective modules are injective.

Show, however, that a ring R is Noetherian iff every direct sum of injective R-modules is injective. (This observation from Bass' graduate student days appears, with reference to Bass, in Chase [1960].)

b. Again, assume that R is a Noetherian ring. Use part a to show that any injective module is a direct sum of indecomposable injective modules.

Exercise A3.3 (Injectives and primes): We shall say that an injective module E is indecomposable if it cannot be written as a direct sum $E = E' \oplus E''$ with both E and E'' nonzero. Suppose that R is a Noetherian ring. Use primary decomposition to show that if E is an indecomposable injective R-module, then $E \cong E(R/P)$ for some prime ideal P of R. Show that if P and Q are primes, then $E(R/P) \cong E(R/Q)$ iff $P = Q$. Thus there is a one-to-one correspondence between indecomposable injectives and prime ideals.

Exercise A3.4: We can compute injective envelopes in some simple cases:

a.* Let R be a Noetherian ring and let P be any ideal of R. Set $E = E(R/P)$. For any ideal $I \subset R$ and map $\varphi : I \to R/P$, use the Artin-Rees lemma (Lemma 5.1) to show that there is a number d such that φ factors through $I/(P^d \cap I) \cong (P^d + I)/P^d$. Deduce that $E' = \cup_d (0 :_E P^d) \subset E$, the set of elements annihilated by some power of P, is injective, and thus that $E = E'$.

b.* With notation as in part a, suppose that P is a maximal ideal. Show that $(0 :_E P^d)$ is the injective hull of R/P over the Artinian ring R/P^d. By Corollary 21.3 it is a module of the same finite length as R/P^d.

c. Let $R = k[x_1, \ldots, x_r]$, and let $P = (x_1, \ldots, x_r)$. Let $E = \oplus_d \operatorname{Hom}_k(R_d, k)$ be the graded dual of R. We have $E \subset E_1 := \operatorname{Hom}_k(R, k) = \Pi_d \operatorname{Hom}_k(R_d, k)$, which is an injective R-module by Lemma A3.8. Show that E is an essential extension of $k = \operatorname{Hom}_k(k, k) \subset \operatorname{Hom}_k(R, k)$. Show that $E = \cup_d (0 :_{E_1} P^d)$. Conclude from part a that E is the injective envelope of k.

d. Show that the indecomposable injective Abelian groups are $\mathbf{Q}$ and, for each prime p, the group

$$\mathbf{Z}/p^\infty := \varinjlim (\mathbf{Z}/p \subset \mathbf{Z}/p^2 \subset \mathbf{Z}/p^3 \subset \cdots) = \mathbf{Z}[p^{-1}]/\mathbf{Z}.$$

Show that $\mathbf{Q}/\mathbf{Z} \cong \oplus_p \mathbf{Z}/p^\infty$.

What is the injective resolution of $\mathbf{Z}/p$ as a $\mathbf{Z}$-module?

Exercise A3.5 (Graded injective modules and injective graded modules): Let $R = \oplus_d R_d$ be a $\mathbf{Z}$-graded ring. If $M = \oplus M_d$ is a graded R-module, and E is an R_0-module, we write

$$\mathrm{Hom}_{\mathrm{gr}}(M, E) := \oplus_d \mathrm{Hom}_{R_0}(M_d, E).$$

This is a graded R-module, and is generally much smaller than $\mathrm{Hom}_{R_0}(M, E) = \Pi_d \mathrm{Hom}_{R_0}(M_d, E)$.

a. Show that if E is an injective R_0-module then $Q = \mathrm{Hom}_{\mathrm{gr}}(R, E)$ is an injective in the category of graded R-modules in the sense that Q satisfies the definition given in the text whenever N and M are graded modules and α is a homomorphism of graded modules. (One way to do this is first to prove an analogue of Lemma A3.4). Conclude that every graded module has a graded-injective resolution.

b. Let $R = k[x]$, where k is a field. Show that $k[x, x^{-1}]$ is injective in the category of graded R-modules. Show that in the category of all R-modules, $k(x)$ is the injective hull of $k[x]$. Conclude that $k[x, x^{-1}]$ is not injective in the category of all R-modules, and that in fact there is no degree-preserving inclusion of R into a graded module that is injective in the category of all modules.

c. Suppose $R = \oplus_{d \geq 0} R_d$ is a positively graded Noetherian ring, and that R_0 is a field. Extend the method of Exercise 3.4c to show that $\mathrm{Hom}_{\mathrm{gr}}(R, k)$ is the injective hull of $k = \mathrm{Hom}_k((R/\oplus_{d>0} R_d), k)$ in the category of all R-modules, not just the category of graded R-modules.

Exercise A3.6 (Injective envelopes and primary decomposition): Still assuming that R is Noetherian, let M be any finitely generated R-module.

a. Let P be a prime. Show that if $\alpha : M \to E(R/P)$ is any map, then $\ker \alpha$ is a P-primary submodule of M.

b.* Show that the injective envelope $E(M)$ is a finite direct sum of indecomposable injectives. Let $M \to E(M) = \oplus E(R/P_i)$ be the injective envelope of M. Show that if P is a prime ideal and if $M(P)$ is the kernel of the composite map $M \to E(M) = \oplus E(R/P_i) \to \oplus_{P_i = P} E(R/P_i)$, then $M(P)$ is P-primary. Show that $0 = \cap M(P)$ is a primary decomposition of 0, and that the set of P that occur among the P_i above is precisely the set $\mathrm{Ass}(M)$.

Exercise A3.7 (More on the Noetherian property): Let $R \subset S$ be rings, and suppose that S is finitely generated as an R-module.

a.* Let F be an R-module. Show that F is injective as an R-module iff $\mathrm{Hom}_R(S, F)$ is injective as an S-module.

b. (Eakin's Theorem) Use part a and the criterion of Exercise A3.2 to show that R is Noetherian iff S is Noetherian. (This result is due to Eakin [1968]; the argument is from Eisenbud [1970]. A direct and more general proof was given by Formanek [1973] and is reproduced in Matsumura [1986, Theorem 3.6].)

A3.5 Basic Constructions with Complexes

A3.5.1 Notation and Definitions

To simplify the notation in what follows, we think of R as a trivially graded ring—that is, the degree $= 0$ part is R and all the other homogeneous components are 0. If

$$F: \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots$$

is a complex, we think of F as a graded R-module (the degree-i component of F is F_i) together with an endomorphism φ of degree -1. As usual we shall make the convention that maps of graded modules have degree 0 unless otherwise specified: Writing $F[i]$ for the graded module obtained from F by the rule $F[i]_j = F_{i+j}$, we could also have said that φ is a map from F to $F[-1]$. Often the grading does not matter, and we define a **differential module** (F, φ) to be an R-module F with an endomorphism φ such that $\varphi^2 = 0$. As for complexes, we define a **cycle** of F to be an element of $\ker \varphi$ and a **boundary** of F to be an element of $\operatorname{im} \varphi$.

Definitions. *Let F be a complex as above. The ith homology module of F is defined to be*

$$H_i(F) = \ker \varphi_i / \operatorname{im} \varphi_{i+1}.$$

We sometimes write $H(F)$ for the direct sum $\oplus_i H_i(F)$ of all the homology modules. If F is simply a differential module, with differential φ, then we set $H(F) = \ker \varphi / \operatorname{im} \varphi$; in case F is a complex, this is again $\oplus_i H_i(F)$.

We say that the complex (or differential module) F is **exact** if $H(F) = 0$. A complex

$$F: \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots \xrightarrow{\varphi_1} F_0$$

is called a **(left) resolution** (of $H_0(F) = \operatorname{coker} \varphi_1$) if $H_i(F) = 0$ for all $i > 0$. (It is sometimes convenient to regard F as continuing to the right forever with 0 maps and modules, thus:

$$F: \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots \xrightarrow{\varphi_1} F_0 \to 0 \to 0 \to \cdots .)$$

If the F_i are projective (respectively, free), then such a resolution is called a **projective (respectively, free) resolution.** Dually, a complex

$$I: I_0 \to I_{-1} \to \cdots I_{-i+1} \to I_{-i} \to I_{-i-1}$$

is called a **(right) resolution** if its only nonzero homology is $H_0(I) = \ker \varphi_0$. A right resolution is called an **injective resolution** if all the I_j are injective modules.

A3.6 Maps and Homotopies of Complexes

Projective (or free) resolutions of modules are in general far from unique (though over a local ring minimal resolutions of finitely generated modules are unique up to noncanonical isomorphisms—see Chapter 20). Thus, if we are to examine modules by studying their resolutions, it is necessary to ask what connects two different resolutions of the same module. This question turns out to have a simple answer. The necessary idea is useful in a more general form.

Definition. *If (F, φ) and (G, ψ) are differential modules, then a* ***map of differential modules*** *is a map of modules $\alpha : F \to G$ such that $\alpha\varphi = \psi\alpha$. If F and G are complexes, then we insist that α preserve the grading as well. Explicitly, if*

$$F: \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots$$

and

$$G: \cdots \to G_{i+1} \xrightarrow{\psi_{i+1}} G_i \xrightarrow{\psi_i} G_{i-1} \to \cdots$$

are complexes of modules, then a map of complexes $\alpha : F \to G$ is a collection of maps

$$\alpha_i : F_i \to G_i$$

of modules making the diagrams

$$\begin{array}{ccccccc} \cdots & \to & F_i & \xrightarrow{\varphi_i} & F_{i-1} & \to & \cdots \\ & & \downarrow{\scriptstyle \alpha_i} & & \downarrow{\scriptstyle \alpha_{i-1}} & & \\ \cdots & \to & G_i & \xrightarrow[\psi_i]{} & G_{i-1} & \to & \cdots \end{array}$$

commutative.

If $\alpha : (F, \varphi) \to (G, \psi)$ is a map of differential modules, then α carries $\ker\varphi$ *to* $\ker\psi$ *and* $\operatorname{im}\varphi$ *to* $\operatorname{im}\psi$. *Thus α gives rise to an* ***induced map on homology,*** *which we also call α:*

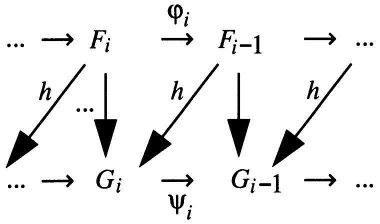
$$\alpha : HF = \frac{\ker\varphi}{\operatorname{im}\varphi} \to \frac{\ker\psi}{\operatorname{im}\psi} = HG.$$

If $\alpha : F \to G$ is a map of complexes, then the grading is preserved, and we get

$$\alpha_i : H_iF = \frac{\ker\varphi_i}{\operatorname{im}\varphi_{i+1}} \to \frac{\ker\psi_i}{\operatorname{im}\psi_{i+1}} = H_iG.$$

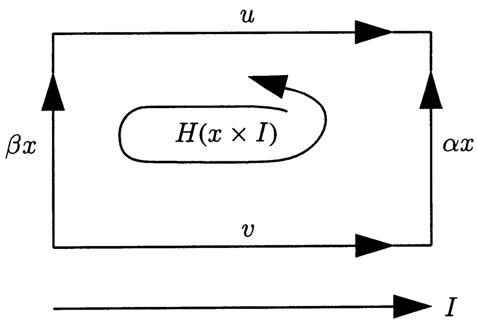
When do two maps of a complex F to a complex G induce the same map on homology? This is a subtle question in general, but there is a very important sufficient condition that may be given in terms of equations. This sufficient condition is called homotopy equivalence.

Definition. *If $\alpha, \beta : (F, \varphi) \to (G, \psi)$ are two maps of differential modules, then α is* **homotopy equivalent** *to β (or simply* **homotopic** *to β) if there is a map of modules $h : F \to G$ such that $\alpha - \beta = \psi h + h\varphi$. If F and G are complexes (so that F and G are graded modules and φ and ψ have degree -1), then we insist that h have degree 1:*



Note that α is homotopy equivalent to β iff $\alpha - \beta$ is equivalent to 0.

The homotopy terminology comes from topology: If α and β are continuous maps from a space X to a space Y, then they induce maps of complexes from the (say, singular) chain complex of X to that of Y. A homotopy $H : X \times I \to Y$ from α to β determines a chain map $h(x) := H(x \times I)$ that raises dimensions by 1. If we orient everything appropriately, we get $\alpha(x) - \beta(x) = \partial(h(x)) - h\partial(x))$ as in Figure A3.2:



$h(x) = H(x \times I);$

$\partial h(x) = \alpha x - u - \beta x + v;$

$h(\delta x) = H(\delta x \times I) = u - v;$

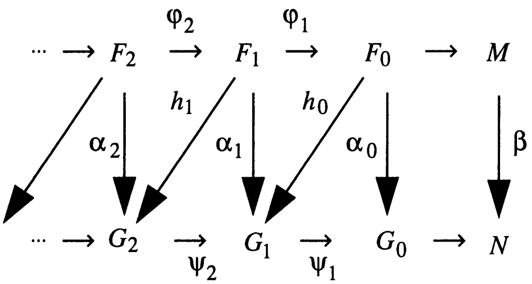
$x\alpha - \beta x = \partial h(x) + h(\partial x).$

FIGURE A3.2.

One of the fundamental properties of homotopic maps in topology is that they induce the same map on homology. The topological proof works by considering the map h on the level of chain complexes. It generalizes immediately to the following algebraic form.

Proposition A3.12. *If $\alpha, \beta : (F, \varphi) \to (G, \psi)$ are two maps of differential modules, and α is homotopy equivalent to β, then α and β induce the same map on homology.*

Proof. It suffices to show that $\alpha - \beta$ induces the map 0 on homology. Thus we may simplify the notation by replacing α by $\alpha - \beta$, and assume from the outset that $\beta = 0$. Let h be the homotopy, so that $\alpha = \psi h + h\varphi$.



Let $x \in \ker \varphi$ be a cycle of F; we must show that $\alpha(x)$ is a boundary of G. From the formula for the homotopy h we get

$$\alpha(x) = \psi(h(x)) + h(\varphi(x)) = \psi(h(x)) + h(0) = \psi(h(x)),$$

as desired. □

An important idea in homological algebra is that one can usefully replace a module with a projective (or dually an injective) resolution. Suppose that F and G are projective resolutions of modules M and N. It turns out that maps from M to N are the same thing as homotopy classes of maps from F to G. An equally useful dual statement, with injective resolutions, can be proved by "dualizing" the following argument; we leave the formulation and proof to the reader.

Proposition A3.13. *Let*

$$F: \;\cdots \to F_i \xrightarrow{\varphi_i} F_{i-1} \cdots \to F_1 \xrightarrow{\varphi_1} F_0$$

and

$$G: \;\cdots \to G_i \xrightarrow{\psi_i} G_{i-1} \cdots \to G_1 \xrightarrow{\psi_1} G_0$$

be complexes of modules, and set $M = \operatorname{coker} \varphi_1 = H_0F$, $N = \operatorname{coker} \psi_1 = H_0G$. *If the modules* F_i *are projective and the homology of* G *vanishes except for* $H_0G = N$, *then every map of modules* $\beta : M \to N$ *is the map induced on* H_0 *by a map of complexes* $\alpha : F \to G$, *and* α *is determined by* β *up to homotopy.*

Proof. Both the existence and the homotopy uniqueness of α are proved by induction; we give the first step and leave the (easy) continuation to the reader.

Existence: Since G_0 maps onto N, the composite map $F_0 \to M \to N$ may be lifted to a map $\alpha_0 : F_0 \to G_0$. It is immediate that $\alpha_0\varphi_1$ maps F_1 to $\ker(G_0 \to N) = \operatorname{im}(G_1 \to G_0)$, so $\alpha_0\varphi_1$ has a lifting $\alpha_1 : F_1 \to G_1$; continuing in this way we get the map of complexes α:

Homotopy uniqueness: If we are given two maps α and α' of complexes lifting the same map $\beta : M \to N$, then subtracting we see that $\alpha - \alpha'$

is a lifting of the zero map. Thus, changing notation, it suffices to show that if α is a lifting of the zero map, then α is homotopic to zero, that is, $\alpha_i = h_{i-1}\varphi_i + \psi_{i+1}h_i$ for some maps $h_i : F_i \to G_{i+1}$. First, since α_0 induces zero : $\operatorname{coker}\varphi_1 \to \operatorname{coker}\psi_1$, it takes F_0 into $\operatorname{im}\psi_1$. Thus there is a lifting $h_0 : F_0 \to G_1$ such that $\psi_1 h_0 = \alpha_0$. Now

$$\psi_1(h_0\varphi_1 - \alpha_1) = \alpha_0\varphi_1 - \psi_1\alpha_1 = 0,$$

so $h_0\varphi_1 - \alpha_1$ maps into $\ker\psi_1 = \operatorname{im}\psi_2$. Since F_1 is projective, we may lift this to a map $h_1 : F_1 \to G_2$. Continuing in this way we get the desired homotopy.

We can at last give the answer to the question with which we began, of what connects different projective resolutions of a module. For later use, we give a version with a functor in it. Recall that a functor F from a category of modules to another category of modules is called **additive** if it preserves the addition of homomorphisms: That is, if $a, b : M \to N$ are homomorphisms, then $F(a+b) = F(a) + F(b) : FM \to FN$. This is the property that we need in order that F preserve homotopy equivalences.

Corollary A3.14.

a. *Any two projective resolutions P and P' of the same module are homotopy equivalent in the sense that there are maps $\alpha : P \to P'$ and $\beta : P' \to P$ such that $\alpha\beta$ is homotopic to the identity map of P' and $\beta\alpha$ is homotopic to the identity map of P.*

b. *If F is any additive functor and we write FP, FP' for the results of applying F to the complexes P and P', then for each i the homology modules $H_i(FP)$ and $H_i(FP')$ are canonically isomorphic.*

Proof.

a. Suppose that P and P' are projective resolutions of a module M. By Proposition A3., there are maps $\alpha : P \to P'$ and $\beta : P' \to P$ of complexes inducing the identity map on M. The composites $\alpha\beta : P' \to P'$ also induces the identity map on M. But the identity map $P' \to P'$ induces the same map on M, so $\alpha\beta$ is homotopic to the identity by the other part of Proposition A3.13. Of course the same argument holds for $\beta\alpha$.

b. Suppose α is as above, and fix an index i. We claim that the map $H_i(F\alpha) : H_i(FP) \to H_i(FP')$ is a canonical isomorphism—that is, an isomorphism independent of the choice of α.

First, if $\alpha' : P \to P'$ were another choice of a map of complexes inducing the identity on M, then by Proposition A3.13 α is homotopic to α', say by a homotopy s with $\alpha - \alpha' = ds + sd$, where d denotes the differential both in P and in P'. Applying F, we get $F\alpha - F\alpha' = FdFs + FsFd$, so $F\alpha$ and $F\alpha'$ induce homotopic maps $FP \to FP'$. By Proposition A3.12, the induced maps $H_i(F\alpha)$ and $H_i(F\alpha')$ are the same.

Next, to see that $H_i(F\alpha)$ is an isomorphism, note simply that $H_i(F\alpha)H_i(F\beta) = H_iF(\alpha\beta) = 1$ because $\alpha\beta$ is homotopic to the identity, and the same argument works for $H_i(F\beta)H_i(F\alpha)$. □

A3.7 Exact Sequences of Complexes

If $\alpha : F' \to F$ and $\beta : F \to F''$ are maps of complexes, with $\beta\alpha = 0$, then we say that

$$0 \to F' \xrightarrow{\alpha} F \xrightarrow{\beta} F'' \to 0$$

is a **short exact sequence of complexes** if for each i the sequence

$$0 \to F_i' \xrightarrow{\alpha_i} F_i \xrightarrow{\beta_i} F_i'' \to 0$$

is exact. Given such a short exact sequence, we get induced maps $\alpha_i : H_iF' \to H_iF$ and $\beta_i : H_iF \to H_iF''$. Somewhat more surprisingly, we get a natural map

$$\delta_i : H_iF'' \to H_{i-1}F'$$

called the **connecting homomorphism**, defined as follows: Write φ', φ, and φ'' for the boundary maps of F', F, and F'', respectively. If $h \in H_iF''$ we choose a cycle $x \in \ker \varphi_i''$ whose homology class is x. Let $y \in F_i$ be an element such that $\beta_i(y) = x$; such a y exists because β_i is surjective. Since $\beta_{i-1}\varphi_i(y) = \varphi_i''\beta_i(y) = \varphi_i''(x) = 0$, there is an element $z \in F_{i-1}'$ such that $\alpha_{i-1}(z) = \varphi_i(y)$. Since α_{i-2} is a monomorphism and $\alpha_{i-2}\varphi_{i-1}'(z) = \varphi_{i-1}\alpha_{i-1}(z) = \varphi_{i-1}\varphi_i(y) = 0$, we see that z is a cycle of F'. We define $\delta_i(h)$ to be the image of z in $H_{i-1}F'$ (see Figure A3.3).

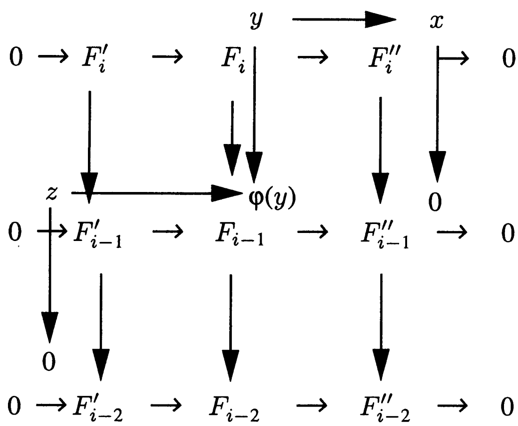


FIGURE A3.3.

A3.7.1 Exercises

Exercise A3.8: Show that $\delta_i(h)$ is independent of the choices made in the definition, and δ_i is a map of modules.

Exercise A3.9: A parallel construction works for exact sequences of differential modules; give it explicitly. The case of complexes becomes a special case if we remark that in the case of complexes the connecting homomorphism can be taken homogeneous and of degree -1.

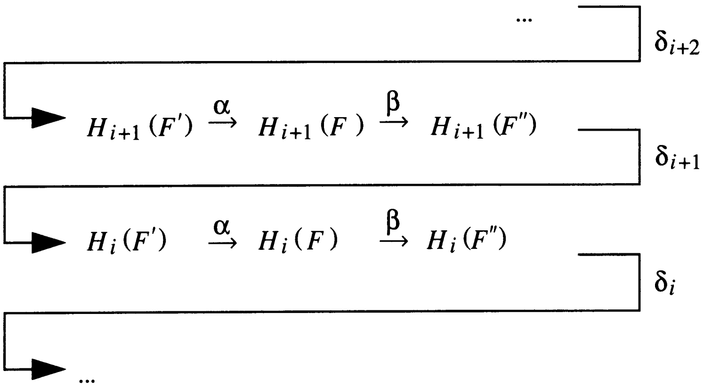


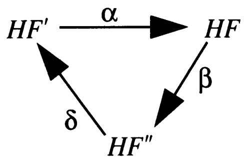
FIGURE A3.4.

A3.8 The Long Exact Sequence in Homology

Proposition A3.15. *If*

$$(*) \qquad 0 \to F' \xrightarrow{\alpha} F \xrightarrow{\beta} F''$$

is a short exact sequence of complexes, then the sequence shown in Figure A3.4, called the **long exact sequence in homology** *of* (∗), *is exact. More generally, if* (∗) *is a short exact sequence of differential modules, then the connecting homomorphism makes the following triangle exact* (*in the sense that the image of each map is the kernel of the next map*).



Proof. We leave the easy verification to the reader. □

Extending the principle embodied in Proposition A3.13, that phenomena regarding modules are well reflected in projective resolutions, we now show that a short exact sequence of modules corresponds to a short exact sequence of projective resolutions in a certain natural sense.

Proposition A3.16. *Let*

$$0 \to M' \xrightarrow{\beta'} M \xrightarrow{\beta} M'' \to 0$$

be a short exact sequence of modules. If

$$F' : \ \cdots \to F_i \xrightarrow{\varphi_i'} F_{i-1} \cdots \to F_1 \xrightarrow{\varphi_1'} F_0,$$

and

$$F'' : \ \cdots \to F_i'' \xrightarrow{\varphi_i''} F_{i-1}'' \cdots \to F_1'' \xrightarrow{\varphi_1''} F_0''$$

are projective resolutions of M' *and* M'', *respectively, then there is a projective resolution* F *of* M *and a short exact sequence of complexes*

$$0 \to F' \xrightarrow{\alpha'} F \xrightarrow{\alpha} F'' \to 0$$

such that α' *and* α *induce the maps* β' *and* β, *respectively.*

Note that because the F_i'' are projective, it follows that $F_i = F_i' \oplus F_i''$ for each i. However, the differentials $\varphi_i : F_i \to F_{i-1}$ of F will generally not be the direct sums of φ_i' and φ_i''.

Proof. Again, we only describe the beginning of the induction, leaving the rest to the reader. Because F_0'' is projective the map from it to M'' can be lifted to a map $F_0'' \to M$. Of course we also have a composite map $F_0' \to M' \to M$. Taking the sum of these maps we get a map $F_0 := F_0' \oplus F_0'' \to M$, and it is easy to check that this is an epimorphism. Replacing M', M, and M'' by the kernels of the maps $F_0' \to M'$, $F_0 \to M$, and $F_0'' \to M''$, respectively, we may repeat this argument. □

A3.8.1 Exercises

Diagrams and Syzygies

Exercises A3.10–A3.12 are three arguments with diagrams that come up so frequently that they have acquired names.

Exercise A3.10 (Snake Lemma):* If

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \end{array}$$

is a commutative diagram of modules with exact rows, show that there is an exact sequence

$$0 \to \ker\alpha \to \ker\beta \to \ker\gamma \to \operatorname{coker}\alpha \to \operatorname{coker}\beta \to \operatorname{coker}\gamma \to 0.$$

Show that if we drop the assumptions that $A \to B$ is a monomorphism and that $B' \to C'$ is an epimorphism, then the six-term sequence is still exact except at the ends.

Where is the "snake"? Look at Figure A3.5.

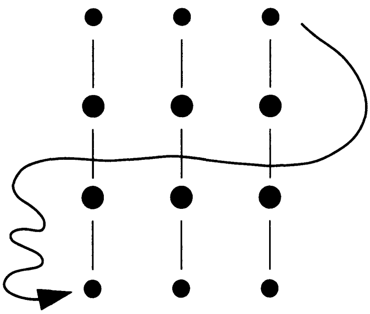


FIGURE A3.5.

Exercise A3.11 (5-Lemma): If

$$\begin{array}{ccccccccc}
A_1 & \longrightarrow & A_2 & \longrightarrow & A_3 & \longrightarrow & A_4 & \longrightarrow & A_5 \\
\alpha_1 \downarrow & & \beta_1 \downarrow & & \gamma \downarrow & & \beta_2 \downarrow & & \alpha_2 \downarrow \\
B_1 & \longrightarrow & B_2 & \longrightarrow & B_3 & \longrightarrow & B_4 & \longrightarrow & B_5
\end{array}$$

is a commutative diagram of modules with exact rows, show that if β_1 and β_2 are isomorphisms, α_1 is an epimorphism, and α_2 is a monomorphism, then γ is an isomorphism. This is often applied when A_1, B_1, A_5, and B_5 are 0.

Exercise A3.12 (9-Lemma): Suppose that the diagram in Figure A3.6

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & A'' & \longrightarrow & B'' & \longrightarrow & C'' & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

FIGURE A3.6.

is a commutative diagram of modules with exact columns, and exact middle row. Show that if either $0 \to A \to B \to C \to 0$ or $0 \to A'' \to B'' \to C'' \to 0$ is exact, then both are.

Exercise A3.13 (Schanuel's Lemma):* Show that if

$$0 \to N_F \to F \to M \to 0$$

and

$$0 \to N_G \to G \to M \to 0$$

are exact sequences with F and G projective, then

$$N_F \oplus G \cong \ker(F \oplus G \to M) \cong N_G \oplus F,$$

where the map in the middle expression is the sum of the two given maps $F \to M$ and $G \to M$.

The module N_F is usually called a **first syzygy module** of M, and its uniqueness "up to projective summand" is another way of saying in what sense the projective resolution of M is unique. (The nth syzygy module is defined inductively as the first syzygy module of the $(n-1)$st syzygy module. Since the first syzygy module of a direct sum may be taken to be the direct sum of the first syzygy modules, all the syzygy modules of M are uniquely defined up to projective summands.)

Exercise A3.14: Let R be a ring and let

$$\mathcal{F}: \ \cdots \to F_d \to \cdots \to F_0 \to M \to 0$$

be a projective resolution of M. Let d be the smallest number such that $\mathrm{im}(F_d \to F_{d-1})$ is projective. Use Schanuel's lemma (Exercise A3.13) to show that d is independent of the resolution chosen, so that $d = \mathrm{pd}\, M$.

A3.9 Derived Functors

One of the main applications of projective and injective resolutions is defining **derived functors**. The idea is this: Often one has a functor F (say, for simplicity, from R-modules to R-modules) that is additive and that takes short exact sequences

$$0 \to A \to B \to C \to 0$$

of modules into sequences that are exact only at one end, say at the right:

$$FA \to FB \to FC \to 0.$$

Such a functor is said to be **right-exact**; an example is the functor $M \otimes_R -$, which takes an R-module to its tensor product with a fixed R-module M. (If the sequence is only exact on the left, we speak of a left-exact functor; an example is $\mathrm{Hom}_R(M, -)$. We shall stick with right-exact functors in the description that follows, and remark on the dualization to the left-exact case at the end. The reader should be warned that we shall apply both notions.)

If F is an interesting right-exact functor, then it is generally interesting to have a description of when a zero can be added on the left end of the right-exact sequence

$$FA \to FB \to FC \to 0$$

and still have an exact sequence; or more generally, to have a good description of the kernel of the left-hand map. Derived functors provide this. In the situation above, for example, there is a "first left-derived functor L_1F" and a map $\delta : L_1F(C) \to FA$ such that

$$L_1F(C) \xrightarrow{\delta} FA \to FB \to FC \to 0$$

is exact. (Here δ must depend on the short exact sequence given, but the module $L_1F(C)$ does not!) Of course, one should then ask about the kernel of δ. In fact, the theory provides a whole sequence of left-derived functors, which answer the sequence of questions beginning in this way:

Definition. *Suppose F is a right-exact functor on the category of R-modules. If A is an R-module, let*

$$P: \quad \cdots \to P_i \xrightarrow{\varphi_i} P_{i-1} \cdots \to P_1 \xrightarrow{\varphi_1} P_0$$

be a projective resolution of A, and define the ***ith left-derived functor*** *of F to be $L_iF(A) = H_iFP$, where FP is the complex*

$$FP: \quad \cdots \to FP_i \xrightarrow{F\varphi_i} FP_{i-1} \cdots \to FP_1 \xrightarrow{F\varphi_1} FP_0,$$

the result of applying F to P.

We have:

Proposition A3.17. *The left-derived functors of F are independent of the choice of resolution and have the following properties:*

a. *$L_0F = F$.*

b. *If A is a projective module, then $L_iF(A) = 0$ for all $i > 0$.*

c. *For every short exact sequence*

$$0 \to A \xrightarrow{u} B \xrightarrow{v} C \to 0,$$

there is a long exact sequence as shown in Figure A3.7.

d. *The "connecting homomorphisms" δ_i in the long exact sequence are* **natural***: That is, if*

$$\begin{array}{ccccccccc} 0 & \to & A & \to & B & \to & C & \to & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \to & A' & \to & B' & \to & C' & \to & 0 \end{array}$$

is a commutative diagram with exact rows (a "map of short exact sequences") then the diagrams

$$\begin{array}{ccc} L_{i+1}FC & \xrightarrow{\delta_{i+1}} & L_iFA \\ \downarrow{\scriptstyle L_{i+1}F\gamma} & & \downarrow{\scriptstyle L_iF\alpha} \\ L_{i+1}FC' & \xrightarrow{\delta_i} & L_iFA' \end{array}$$

commute.

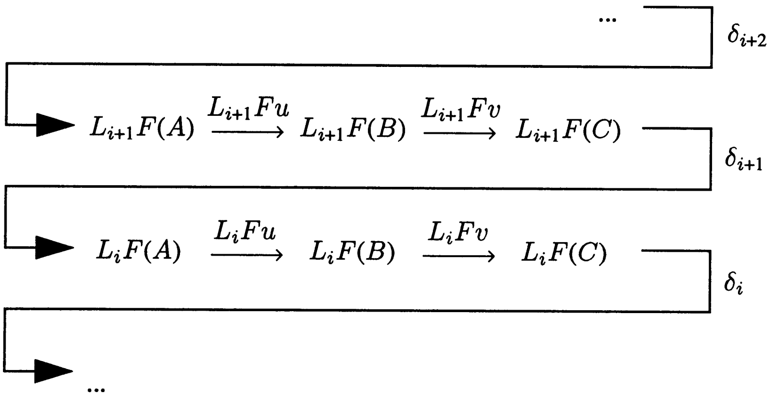


FIGURE A3.7.

Proof. The independence of resolution is the content of Corollary A3.14b.

a. To show that $L_0F(A) = F(A)$, just use the right-exactness of F: From the definition

$$L_0F(A) = H_0(\cdots \to FP_1 \to FP_0),$$

we get $L_0F(A) = \operatorname{coker} FP_1 \to FP_0 = FA$.

b. This is immediate from the independence of resolution, since if A is projective then we may take as projective resolution the complex

$$\cdots \to 0 \to 0 \to A.$$

c. This is immediate from Propositions A3.15 and A3.16.

d. Form the projective resolutions of each of the two short exact sequences as in Proposition A3.16. The maps α, β, and γ lift to comparison maps between these resolutions. If we use these maps of resolutions to define the maps $L_iF(\alpha)$ and $L_iF(\beta)$, then the verification of the commutativity of the diagram in part d is easy. We leave the details to the reader. □

Dually, if F is a left-exact functor, then we define the right-derived functors R^iF of F: If A is a module, we let

$$Q : 0 \to Q_0 \to Q_{-1} \to \cdots$$

be an injective resolution of A, and we set

$$R^iF(A) = H_{-i}(FQ),$$

where FQ is the complex

$$FQ : 0 \to FQ_0 \to FQ_{-1} \to \cdots .$$

Proposition A3.17 dualizes to this setting.

A3.9.1 Exercise on Derived Functors

Exercise A3.15:* Show that the conditions of Proposition A3.17 characterize the functors L_iF.

A3.10 Tor

Let A be an R-module. The left-derived functors of the functor $M \otimes_R -$ are called $\mathrm{Tor}_i^R(M, -)$. The tensor product itself is commutative in the sense that $M \otimes_R N \cong N \otimes_R M$, and this property carries over to the Tor_i, as we shall prove in the section on spectral sequences. Thus Tor_i^R may be regarded as a functor of two variables, $\mathrm{Tor}_i^R(-, -)$, and we get long exact sequences from short exact sequences in either variable. When the ring R is understood, we suppress it from the notation. We give a few very useful computations as exercises; the reader is urged to do at least the first three.

A3.10.1 Exercises: Tor

The name "Tor" comes from the following computation, which connects Tor with torsion.

Exercise A3.16:* Let $x \in R$ be a nonzerodivisor. Show that

$$\mathrm{Tor}_1(R/x, M) = \{m \in M | xm = 0.\}$$

Exercise A3.17: If I and J are any ideals of R, then $IJ \subset I \cap J$. Show that $\mathrm{Tor}_1(R/I, R/J) = (I \cap J)/(IJ)$. This usefully encapsulates several often-used cases (of course these can also be proven directly). For example, use it to show that $I \cap J = IJ$ in the following cases:

a. $I + J = R$.

b. I is generated by a sequence of elements that form a regular sequence mod J.

Exercise A3.18 ("Betti" numbers): Let $(R, \mathfrak{m})$ be a local ring. We say that a free resolution

$$F : \cdots \to F_{i+1} \xrightarrow{\varphi_{i+1}} F_i \xrightarrow{\varphi_i} F_{i-1} \to \cdots \xrightarrow{\varphi_1} F_0$$

of a module M is **minimal** if each φ_i has an image contained in $\mathfrak{m}F_{i-1}$. (If the F_i are finitely generated modules, then Nakayama's lemma shows that this is equivalent to a more obviously natural formulation. See Chapter 20.) If F as above is a minimal free resolution of M and rank $F_i = b_i$, then show

that $\operatorname{Tor}_i(R/\mathfrak{m}, M) = (R/\mathfrak{m})^{b_i}$. The b_i are called Betti numbers of M, in loose analogy with the situation in topology, where F is a chain complex.

Exercise A3.19 (Serre's Intersection Formula): Let X and Y be subvarieties of $\mathbf{A}^r_k$, of dimensions d and $n-d$, defined by ideals I and $J \subset k[x_1, \ldots, x_r] = S$, and suppose that $X \cap Y$ has the origin 0 as an isolated point. A crucial part of algebraic geometry is devoted to the question, in this and similar cases, of defining an "intersection multiplicity $i(X, Y; 0)$ of X and Y at 0" that will have desirable properties. If X and Y are themselves nice (for example, nonsingular at 0), then this is not too hard; writing R for the localization of S at $(x_1, \ldots, x_r)$, the right answer turns out to be the vector space dimension of $R/I \otimes_R R/J = R/(I+J)$. Such a formula is correct also in the case of plane curves, but in general the dimension of $R/(I+J)$ turns out only to be the first term of an alternating sum. The following definition is due to Serre [1957]:

$$i(X, Y; 0) := \sum_j (-1)^j \dim_k \operatorname{Tor}_j^R(R/I, R/J).$$

Show that $\operatorname{Tor}_j^R(R/I, R/J)$ is annihilated by both I and J, and therefore has finite length. Let $r = 4$, and take $I = (x_1, x_2) \cap (x_3, x_4)$, the ideal corresponding to the union X of two two-planes, meeting in the point 0, and $J = (x_1 - x_3, x_2 - x_4)$, the ideal corresponding to another two-plane Y, transverse to each of the first two and meeting them at the origin. Compute the $\operatorname{Tor}_j^R(R/I, R/J)$ and show that $i(X, Y; 0) = 2$. Note that Y meets each of the two-planes in X transversely in a single point (multiplicity 1) so Y "should" meet X with multiplicity 2; however, the length of $R/I \otimes_R R/J = R/(I+J)$ is not 2.

Exercise A3.20 (Tor as an algebra): For any R-modules A, A', B, B', define a natural "external multiplication" map

$$e : \operatorname{Tor}_m^R(A, B) \otimes_R \operatorname{Tor}_n^R(A', B') \to \operatorname{Tor}_{m+n}^R(A \otimes_R A', B \otimes_R B')$$

as follows. Let P and P' be projective resolutions of A and A'. Represent elements α, β of $\operatorname{Tor}_m^R(A, B)$ and $\operatorname{Tor}_n^R(A', B')$ as cycles of the complexes $P \otimes B$ and $P' \otimes B'$ (where for simplicity we write $\otimes$ for $\otimes_R$). Show that $\alpha \otimes \beta$ is then naturally a cycle in the tensor product complex $(P \otimes B) \otimes (P' \otimes B') \cong P \otimes P' \otimes B \otimes B'$. (Here the tensor product of two complexes may be defined as the total complex of the double complex with terms $P_i \otimes B \otimes P'_j \otimes B'$—see the section on double complexes below if this is unfamiliar.) If P'' is a free resolution of $A \otimes A'$, there is a map of complexes $P \otimes P' \to P''$ inducing the identity on $H_0 = A \otimes A'$. Use this to define e.

If A and B are R-algebras, take $A' = A$ and $B' = B$ and combine the map above with the multiplication maps of A and B to get a multiplication

$$\mu : \operatorname{Tor}_m^R(A,B) \otimes_R \operatorname{Tor}_n^R(A,B) \to \operatorname{Tor}_{m+n}^R(A,B).$$

Show that this makes $\operatorname{Tor}_*^R(A,B)$ into a graded associative R-algebra that is "graded-commutative" in the sense that for elements α, β of degrees a and b we have

$$\beta\alpha = (-1)^{ab}\alpha\beta.$$

Remarks: A good deal of work has been done on the structure of this algebra in the case where $A = B = k$, the residue class field of a local ring R. In that case Tate and Gulliksen showed, for example, that $\operatorname{Tor}_*^R(k,k)$ is a free graded-commutative divided power algebra (that is, the tensor product of a divided power algebra on even degree generators and an exterior algebra on odd degree generators). It was hoped for a long time that the "Poincaré series" of R, namely the power series

$$P_R(t) = \sum_n \dim_k(\operatorname{Tor}_n^R(k,k))t^n,$$

would be a rational function of t, but Anick [1982] showed that this is false in general. The hope behind this hope was perhaps that the ranks of the free modules in a minimal free resolution of k are "finitely determined." It remains an open problem to give a description simpler than the one obtained by computing the minimal free resolution.

One important point in this development is that the algebra structure on Tor can be computed from a resolution that is an algebra in a nice way:

Exercise A3.21: Let R be a ring with augmentation onto a factor ring $R \to k$. Suppose that

$$P : \cdots \xrightarrow{d} P_1 \xrightarrow{d} P_0$$

is a projective resolution of k over R, with $P_0 = R$. Suppose that the complex P has an algebra structure such that d is a derivation, $d(pq) = d(p)q + (-1)^q pd(q)$. Show that this algebra structure induces the natural algebra structure on the homology $\operatorname{Tor}(k,k) = H_*(P \otimes k)$.

Exercise A3.22 (Auslander's Transpose Functor): The long exact sequence in Tor is not the only answer to the question of how to measure the inexactness of the functor $\otimes$. Suppose that M is a finitely presented R-module. Following ideas of Auslander [1966], we define the **transpose** of M as follows:

Let $\varphi : F \to G$ be a **projective presentation of** M—that is, a map of projective modules with $\operatorname{coker} \varphi = M$. Write $-^*$ for $\operatorname{Hom}_R(-,R)$, so that $\varphi^* : G^* \to F^*$ is the "transpose" of φ. Define $T(\varphi)$, the transpose of M, to be $T(\varphi) = \operatorname{coker} \varphi^*$.

a.* Show that like the first syzygy of M, $T(\varphi)$ depends, up to a projective summand, only on M in the sense that if φ' is another projective presentation of M, then there are projective modules P and P' such that $T(\varphi) \oplus P' \cong T(\varphi') \oplus P$.

Notation: We shall write $T(M)$ for any (fixed) choice $T(\varphi)$. We may choose things so that $T(T(M)) = M$.

b. Show that if

$$\alpha : 0 \to A \to B \to C \to 0$$

is a short exact sequence of R-modules, and M is a finitely presented R-module, then there is an exact sequence

$$0 \to \operatorname{Hom}(T(M), A) \to \operatorname{Hom}(T(M), B) \to \operatorname{Hom}(T(M), C) \to M \otimes A \to M \otimes B \to M \otimes C \to 0.$$

This sequence gives another way of "measuring" the inexactness of the functor $M \otimes -$. If N is any module, and we choose $M = T(N)$, then since $T(T(N)) = N$, we may also think of it as a measure for the inexactness of $\operatorname{Hom}(N, -)$.

c. Here is an application: We say that $A \subset B$ is a **pure** R-submodule if for every module M the induced map $M \otimes_R A \to M \otimes_R B$ is a monomorphism. Show that if $\alpha : 0 \to A \to B \to C \to 0$ is a short exact sequence with $A \to B$ pure, and if N is a finitely presented R-module, then

$$\operatorname{Hom}(N, \alpha) : 0 \to \operatorname{Hom}(N, A) \to \operatorname{Hom}(N, B) \to \operatorname{Hom}(N, C) \to 0$$

is a short exact sequence. Deduce that if C is finitely presented, then α splits. Note that it is even enough to know that $N \otimes A \to N \otimes B$ is a monomorphism for every finitely presented module N. (This is actually the same as purity, since every module is the filtered direct limit of finitely presented modules—see Exercise A6.5.)

A3.11 Ext

We now turn from $\otimes$ to Hom. The functor $\operatorname{Hom}_R(M, -)$ is left-exact, so we may apply the dual theory, the theory of right-derived functors, as follows: For any R-module N, let

$$I : I_0 \to I_1 \to \cdots$$

be an injective resolution of N, and define the right-derived functor $R^i \operatorname{Hom}(M, -)(N)$, which we shall write more compactly as $\operatorname{Ext}^i_R(M, N)$, to be $H_{-i}(\operatorname{Hom}_R(M, I))$, where $\operatorname{Hom}_R(M, I)$ is the complex

$$\operatorname{Hom}_R(M, I) : 0 \to \operatorname{Hom}_R(M, I_0) \to \operatorname{Hom}_R(M, I_1) \to \cdots .$$

As we shall prove by spectral sequences later (another proof, done by identifying both results with the "Yoneda Ext," is given in the exercises), we could also compute this from a projective resolution

$$F : \cdots \to F_1 \to F_0$$

of M as $\operatorname{Ext}^i_R(M, N) = H_{-i}(\operatorname{Hom}_R(F, N))$, where $\operatorname{Hom}_R(F, N)$ is the complex

$$\operatorname{Hom}_R(F, N) : 0 \to \operatorname{Hom}_R(F_0, N) \to \operatorname{Hom}_R(F_1, N) \to \cdots .$$

Here is a classic application of Ext, due to Auslander, showing that the global dimension of a ring can be computed from finitely generated modules—even from cyclic modules. The original proof used a direct limit argument; the proof given here, using injective modules, is due to Serre. The result is very general: It holds for non-Noetherian rings too, and even for noncommutative rings if we specify left or right modules and ideals throughout.

Theorem A3.18 (Auslander [1955]). *The following conditions on a ring R are equivalent:*

a. $\operatorname{gl\,dim} R \le n$*—that is,* $\operatorname{pd} M \le n$ *for every R-module M.*

b. $\operatorname{pd} R/I \le n$ *for every ideal I.*

c. injective dimension $N \le n$ for every R-module N.

d. $\operatorname{Ext}^i_R(M, N) = 0$ *for all $i > n$ and all R-modules M and N.*

Proof.

a $\Rightarrow$ b is trivial.

b $\Rightarrow$ c: Suppose that condition b holds and let

$$0 \to N \to E_0 \to \cdots \to E_{n-1} \to X \to 0$$

be an exact sequence with the E_i injective; we shall show that X is injective, proving c. Breaking the long exact sequence above into short exact sequences, and considering the long exact

sequences obtained from these by applying $\mathrm{Ext}^*_R(R/I, -)$, we see that

$$\mathrm{Ext}^1_R(R/I, X) \cong \mathrm{Ext}^{n+1}_R(R/I, M) = 0,$$

the last equality coming from the hypothesis b. Thus it suffices to show that a module X is injective if $\mathrm{Ext}^1_R(R/I, X) = 0$ for all ideals I. Computing $\mathrm{Ext}^1_R(R/I, X)$ from a projective resolution of R/I, we see that this hypothesis is equivalent to saying that if $\psi : I \to X$ is any map, then there is a map $R \to X$ such that the composition $I \to R \to X$ is ψ. By Lemma A3.4, X is injective.

c $\Rightarrow$ d: Compute $\mathrm{Ext}_R(M, N)$ from an injective resolution of N.

d $\Rightarrow$ a: Assume that condition d holds, and let

$$0 \to X \to F_n \to \cdots \to F_0 \to M \to 0$$

be an exact sequence with the F_i projective. It will suffice to show that X is projective. Applying the long exact sequences in Ext to the short exact sequences

$$0 \to \ker(F_{i+1} \to F_i) \to F_i \to \ker(F_i \to F_{i-1}) \to 0$$

obtained from this resolution, we see that

$$\mathrm{Ext}^1_R(X, N) \cong \mathrm{Ext}^{n+1}_R(M, N) = 0,$$

for every module N. We shall show that this condition implies that X is projective. (Note the duality of this with the preceding argument —but here there is no restriction on N, and the proof is easier.)

To this end we must show that if

$$\mathcal{P} \cdots \to P_1 \to P_0 \to X \to 0$$

is a projective resolution, then the map $P_0 \to X$ splits. Let

$$N = \ker(P_0 \to X).$$

The natural map $\varphi : P_1 \to N$ is a *cycle* of $\mathrm{Hom}(\mathcal{P}, N)$ and thus defines an element of $\mathrm{Ext}^1_R(X, N)$; since this group vanishes, the element is a boundary so there exists a map $P_0 \to N$ extending φ. This map is a splitting of the inclusion $N \to P_0$, and thus coker $P_0 \to X$ splits too. This concludes the proof of the equivalence of conditions a–d. □

As with Tor, we offer the reader some simple exercises to become comfortable with Ext.

A3.11.1 Exercises: Ext

Exercise A3.23: If x is a nonzerodivisor in a ring R, compute $\operatorname{Ext}^i_R(R/x, M)$. In particular, compute $\operatorname{Ext}^i_{\mathbf{Z}}(\mathbf{Z}/n, \mathbf{Z}/m)$ for any integers n, m.

Exercise A3.24: Show that a finitely generated Abelian group A is free iff $\operatorname{Ext}^1_{\mathbf{Z}}(A, \mathbf{Z}) = 0$. It was conjectured by Whitehead that this would hold for all groups, but the truth turns out to depend on your set theory (Shelah [1974]).

Exercise A3.25:* For any ring R and ideal $I \subset R$, show from the definitions and Exercise A3.17 that

$$\operatorname{Ext}^1_R(R/I, R/I) = \operatorname{Hom}_R(I/I^2, R/I) = \operatorname{Hom}(\operatorname{Tor}_1(R/I, R/I), R/I).$$

In a geometric context, supposing that R is the affine coordinate ring of a variety X and that I is the ideal of a subvariety Y, this module $\operatorname{Hom}_R(I/I^2, R/I)$ plays the role of the "normal bundle" of Y in X; see Exercise 16.8 for more information.

Exercise A3.26 (Yoneda's description of $\mathbf{Ext}^1$): The ideas in this and the next exercise give a useful and appealing interpretation of the elements of Ext. See, for example, MacLane [1963, Chapter III] for more details.

a. If

$$\begin{aligned} \alpha: &\quad 0 \to B \to X \to A \to 0 \\ \alpha': &\quad 0 \to B \to X' \to A \to 0 \end{aligned}$$

are short exact sequences, we say that α is **Yoneda equivalent** to α' if there exists a map $f: X \to X'$ making the diagram

$$\begin{array}{ccccccccc} 0 & \to & B & \to & X & \to & A & \to & 0 \\ & & \| & & {\scriptstyle f}\downarrow & & \| & & \\ 0 & \to & B & \to & X' & \to & A & \to & 0 \end{array}$$

commute. Show that Yoneda equivalence is an equivalence relation (reflexive, symmetric, and transitive). Show that α is Yoneda equivalent to the "split" sequence

$$0: 0 \to A \to A \oplus B \to B \to 0$$

iff α is itself split.

We shall write $[\alpha]$ for the Yoneda equivalence class of a short exact sequence α.

We now define $E^1(A, B)$ to be the set of equivalence classes of short exact sequences as above. We shall see that $E^1_R(A, B)$ is naturally isomorphic to $\operatorname{Ext}^1_R(A, B)$.

b. Functoriality in A: Show that $E^1_R(A,B)$ is a contravariant functor of A as follows: If

$$\alpha : 0 \to B \to X \xrightarrow{a} A \to 0$$

is a short exact sequence and $v : A' \to A$ is a map, define

$$X' = \ker(-a, v) : X \oplus A' \to A,$$

X' is called the **pull-back (or fibered product)** of X and A' over A. Show that there is a short exact sequence

$$\alpha' : 0 \to B \to X' \to A' \to 0$$

and a commutative diagram

$$\begin{array}{ccccccccccc} \alpha': & 0 & \to & B & \to & X' & \to & A' & \to & 0 \\ & & & \| & & \downarrow & & \downarrow v & & \\ \alpha: & 0 & \to & B & \to & X & \to & A & \to & 0. \end{array}$$

We define $v([\alpha])$ to be α'. Show that this makes $E^1_R(A,B)$ into a contravariant functor of A as claimed.

c. Functoriality in B: Given a map $b : B \to X$ and another map $u : B \to B'$, the **push-out (or fibered coproduct)** of X and B' under B is by definition $\operatorname{coker}(-b, u) : B \to X \oplus B'$. Dualize the argument of part b, using the push-out construction, to show that $E^1_R(A,B)$ is a covariant functor of A.

d. Prove that $E^1_R(A,B) \cong \operatorname{Ext}^1_R(A,B)$ as follows: Let

$$Q : Q_0 \xrightarrow{\psi_0} Q_{-1} \xrightarrow{\psi_{-1}} \cdots$$

be an injective resolution of B, and let $b : B \to Q_0$ be the injection of B to Q_0 that is the kernel of ψ_0. An element ν of $\operatorname{Ext}^1_R(A,B)$ is represented by a cycle of $\operatorname{Hom}_R(A,Q)$, which is a map $v : A \to Q_{-1}$ such that $\psi_{-1}v = 0$; that is, a map $v : A \to \ker \psi_{-1} = Q_0/B$. Let α be the short exact sequence

$$\alpha : 0 \to B \to Q_0 \to Q_0/B \to 0,$$

and let $\nu' \in E^1_R(A,B)$ be the element $v([\alpha])$. Show that

$$\varepsilon : \operatorname{Ext}^1_R(A,B) \to E^1_R(A,B); \nu \mapsto \varepsilon(\nu) := \nu'$$

is a bijection of sets, natural in the sense that if $A' \to A$ or $B \to B'$ are homomorphisms, then the induced maps on $\operatorname{Ext}^1_R(A,B)$ and $E^1_R(A,B)$ correspond. If $P : \cdots \to P_1 \to P_0$ is a projective resolution of A, show dually that $E^1_R(A,B)$ may be identified with $H_{-1}(\operatorname{Hom}_R(P,B))$. This proves that $\operatorname{Ext}^1_R(A,B)$ could be computed from a projective resolution of A as well as from an injective resolution of B.

e. The module structure on E^1_R : If $r \in R$, the underlying ring, then multiplication by r is an endomorphism of any module B, and thus induces a map on $E^1_R(A,B)$ by functoriality in B. Of course, it also induces a map by functoriality in A. Show that these two maps are the same; we use them to define the action of R on $E^1_R(A,B)$. To define an addition on $E^1_R(A,B)$, let α, α' be short exact sequences as in part a. Let $d : A \to A \oplus A$ be the diagonal map $d(a) = (a,a)$, and let $s : B \oplus B \to B$ be the sum map $s(b,b') = b + b'$. Let $\alpha \oplus \alpha'$ be the direct sum of α and α',

$$\alpha \oplus \alpha' : 0 \to B \oplus B \to X \oplus X' \to A \oplus A \to 0$$

and set

$$[\alpha] + [\alpha'] = sd[\alpha \oplus \alpha'] = ds[\alpha \oplus \alpha'].$$

Show that these definitions make $E^1_R(A,B)$ a module and ε an isomorphism of modules.

f. If

$$\beta : 0 \to B' \to B \to B'' \to 0$$

is a short exact sequence of modules, define a "connecting homomorphism" $\delta : \mathrm{Hom}_R(A,B'') \to E^1_R(A,B')$ for $b \in \mathrm{Hom}_R(A,B'')$ by $\delta(b) = b[\beta] \in E^1_R(A,B')$. Show that there is an exact sequence

$$\begin{aligned} 0 \to \mathrm{Hom}_R(A,B') &\to \mathrm{Hom}_R(A,B) \to \mathrm{Hom}_R(A,B'') \\ &\xrightarrow{\delta} E^1_R(A,B') \to E^1_R(A,B) \to E^1_R(A,B''), \end{aligned}$$

and that if we identify E^1_R with Ext^1_R, then δ is the usual connecting homomorphism.

Exercise A3.27 (Ext as an algebra; the Yoneda Ext in general):

a. Higher Exts: Two exact sequences from A to B "of length n"

$$\begin{aligned} \alpha :\ & 0 \to A \to X_1 \to X_2 \to \cdots \to X_n \to B \to 0 \\ \alpha' :\ & 0 \to A \to X'_1 \to X'_2 \to \cdots \to X'_n \to B \to 0 \end{aligned}$$

are **primitively equivalent** if there is a commutative diagram

$$\begin{array}{ccccccccccccccc} \alpha : & 0 & \to & A & \to & X_1 & \to & X_2 & \to & \cdots & \to & X_n & \to & B & \to & 0 \\ & & & \| & & \downarrow & & \downarrow & & \cdots & & \downarrow & & \| & & \\ \alpha' : & 0 & \to & A & \to & X'_1 & \to & X'_2 & \to & \cdots & \to & X'_n & \to & B & \to & 0. \end{array}$$

This is not an equivalence relation (it is not symmetric), but we may define Yoneda equivalence to be the equivalence relation it generates. Define $E^n_R(A,B)$ to be the set of Yoneda equivalence classes of exact sequences of length n from A to B. Analogously with the case done in the previous exercise, show that $E^n_R(A,B)$ is naturally isomorphic to $\mathrm{Ext}^n_R(A,B)$ (computed from either an injective resolution of B or a projective resolution of A).

b. The Yoneda product: The functoriality of $\operatorname{Ext}^1(A,B)$ may be thought of as giving rise to "multiplication" maps

$$\begin{aligned}\mu: \quad \operatorname{Hom}_R(B,C)\otimes_R \operatorname{Ext}^n_R(A,B) &\to \operatorname{Ext}^n_R(A,C);\\ \mu: \quad \operatorname{Ext}^m_R(B,C)\otimes_R \operatorname{Hom}_R(A,B) &\to \operatorname{Ext}^m_R(A,C).\end{aligned}$$

Thinking of Hom as Ext^0 is the first step in defining an "algebra structure," which is a natural pairing called the Yoneda product

$$\mu: \operatorname{Ext}^n_R(B,C)\otimes_R \operatorname{Ext}^m_R(A,B) \to \operatorname{Ext}^{n+m}_R(A,C)$$

defined for all m and n. Namely, if

$$\alpha: \ 0\to A\to X_1\to X_2\to\cdots\to X_m \xrightarrow{b} B\to 0$$

and

$$\beta: \ 0\to B\xrightarrow{b'} Y_1\to Y_2\to\cdots\to Y_n\to C\to 0$$

are exact sequences, then we define $\mu([\beta]\otimes[\alpha])$ to be the class of the exact sequence

$$\beta\alpha: \ 0\to A\to X_1\to\cdots\to X_m\xrightarrow{b'b} Y_1\to\cdots\to Y_n\to C\to 0.$$

Prove that this multiplication is well defined on Yoneda equivalence classes and that it is associative. (The only case that needs work is where one of the factors is in $\operatorname{Ext}^0 = \operatorname{Hom}$.)

Note that the Yoneda algebra defined above is graded by the positive integers and pairs of modules! However, if we fix a module A and take $A = B$, we get a more reasonable object, a (noncommutative) algebra $\operatorname{Ext}_R(A,A) := \oplus_{n\geq 0}\operatorname{Ext}^n_R(A,A)$ that is graded by the positive integers. Very little is known in general about the properties of this algebra, although extensive work has been done on the case where R is local and $A = k$ is its residue class field. The natural commutative algebra structure on $\operatorname{Tor}^R(k,k) := \oplus_n \operatorname{Tor}^R_n(k,k) = \oplus_n \operatorname{Hom}_k(\operatorname{Ext}^n_R(k,k),k)$, described in the exercises on Tor, makes $\operatorname{Ext}^n_R(k,k)$ into a cocommutative Hopf algebra. Good references are Gulliksen-Levin [1969] for the early work and Anick [1988] for more recent developments. One important point (used in the exercises of Chapter 17) is that the product on $\operatorname{Ext}(k,k)$ may be computed from an appropriate coalgebra structure of the resolution of k.

Exercise A3.28:* Let R be a ring with augmentation onto a factor ring $R\to k$. Suppose that

$$P: \ \cdots \xrightarrow{d} P_1 \xrightarrow{d} P_0$$

is a projective resolution of k over R, with $P_0 = R$. Suppose that the complex $P^* = \operatorname{Hom}_R(P,R)$ has the structure of a graded algebra over R

and that the differential d^* of P^* is defined by multiplication by some fixed element $x \in P_1$, that is $d^*(p) = xp$. Show that the algebra structure on P^* induces the Yoneda algebra structure on $\mathrm{Ext}_R(k, k)$, in the sense that the cycles of P^* form a subalgebra of P^* and the map from the cycles onto $\mathrm{Ext}_R(k, k)$ is an algebra homomorphism. Show that if R is a regular local ring then the hypothesis on d^* is satisfied.

Exercise A3.29 (Miyata [1967]):

a.* (Apparently split implies split) If $\alpha : 0 \to A \to B \to C \to 0$ is a short exact sequence of finitely generated modules over a Noetherian ring and $B \cong A \oplus C$, then α splits. If you find the general case difficult, try the case where A, B, and C are finite Abelian groups.

b. One could try to classify the R-modules X that are extensions of one given module B by another, A, by classifying elements of $\mathrm{Ext}^1_R(A, B)$. One problem with this approach is that one can have two short exact sequences

$$\alpha : 0 \to B \to X \to A \to 0$$

and

$$\alpha' : 0 \to B \to X' \to A \to 0$$

with $X \cong X'$ without α being Yoneda-equivalent to α'. Give an example of sequences of finite Abelian groups where this happens. In general, it is hard to say even what relationship $[\alpha]$ and $[\alpha'] \in \mathrm{Ext}^1_R(A, B)$ have. However, part a shows that $[\alpha] = 0$ iff $[\alpha'] = 0$. Extend this by proving, with notation as above, that if $X \cong X'$, then

$$\mathrm{rad}(\mathrm{ann}[\alpha]) = \mathrm{rad}(\mathrm{ann}[\alpha']).$$

A3.11.2 Local Cohomology

The third derived functor of great use in commutative algebra is **local cohomology**. (The coherent sheaf cohomology of the algebraic geometers can also be expressed in terms of it, at least for projective varieties, and local cohomology with I the ideal of a subvariety plays in a certain sense the role of "relative" cohomology; see Grothendieck [1967].) For any ideal I of R, let $\Gamma_I(M) = \{m \in M | I^p m = 0 \text{ for sufficiently large } p\}$. It is easy to see that Γ_I is a left-exact functor, and we define

$$H^i_I(M) = R^i\Gamma_I(M),$$

again as the homology of the complex obtained by applying Γ_I to an injective resolution of M. We explain something of the properties of this derived functor in the central case where R is a local ring and I is the maximal ideal in Appendix 4.

Part II: From Mapping Cones to Spectral Sequences

A3.12 The Mapping Cone and Double Complexes

If $\alpha : F \to G$ is a map of complexes, then in many contexts we would like to know about the kernel and cokernel of the map induced by α on homology. If α were part of a short exact sequence of complexes—that is, if either all the α_i were monomorphisms or all were epimorphisms, then we could study this problem by looking at the corresponding long exact sequence in homology. Of more general usefulness is the following simple way of producing an exact sequence of complexes

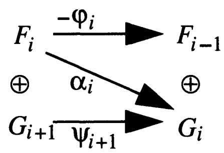
$$0 \to G \to M \to F[-1] \to 0$$

whose connecting homomorphisms are the maps on homology

$$\alpha_i : H_{i-1}F = H_i(F[-1]) \to H_{i-1}G$$

induced by α. Here we make the convention that if F is a complex with differential φ, then $F[i]$ is the complex where $F[i]_j = F_{i+j}$ and with differential $(-1)^i\varphi$. Of course the change of sign of the differential has no effect on the homology module (indeed, the complexes with signs changed or not are isomorphic—the map is -1 in every degree), but turns out to be convenient.

Definition. *If $\alpha : F \to G$ is a map of complexes, and we write φ and ψ, respectively, for the differentials of F and G, then the* **mapping cone** *$M(\alpha)$ of α is the complex such that $M(\alpha)_i = F_{i-1} \oplus G_i$, with differential*



That is, on G_{i+1} the map is the differential of G, but on F_i the map is the sum of the differential of F and the given map α of complexes.

Again, the motivation for this construction is topological: If $\alpha : X \to Y$ is a continuous map between topological spaces, then we may form the union $X \times I \cup Y$. Let M be the space obtained by identifying $X \times \{0\}$ to a point, and $X \times \{1\}$ to $\alpha(X)$ in Y, as in Figure A3.8. The d-dimensional chains of M are generated by the d-dimensional chains of Y and in addition for every (d-1)-chain x of X, a d-dimensional chain that we may describe as

$$\frac{\tilde{x} := (x \times I) \cup \alpha(x)}{x \times \{0\} = \text{point}, x \times \{1\} = \alpha(x) \subset Y}.$$

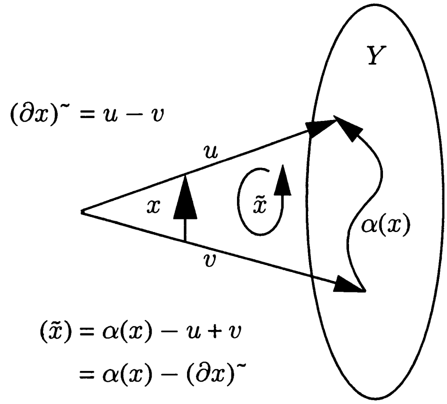


FIGURE A3.8.

With orientations as in Figure A3.8, we have $\partial(\tilde{x}) = -(\partial(x))^{\sim} + \alpha(x)$.

Proposition A3.19. *With notation as in the preceding definition, the natural inclusion makes G into a subcomplex of $M(\alpha)$, and $M(\alpha)/G \cong F[-1]$, so that there is a short exact sequence*

$$0 \to G \to M(\alpha) \to F[-1] \to 0$$

of complexes. In the corresponding long exact sequence in homology,

$$\cdots \to H_i(G) \to H_iM(\alpha) \to H_i(F[-1]) \xrightarrow{\delta} H_{i-1}(G) \to \cdots,$$

the connecting homomorphism δ is the map $H_i(F[-1]) = H_{i-1}F \to H_{i-1}G$ induced on homology by $\alpha : F \to G$.

Proof. The fact that G is a subcomplex of $M(\alpha)$ with quotient $F[-1]$ (that is, F shifted by 1 in degree) follows at once from the definition. To compute the effect of the connecting homomorphism, recall that if $[z]$ is the homology class in $H_i(F[-1])$ of a cycle z of degree i, then $\delta([z])$ is by definition the homology class of $d\tilde{z}$, where $\tilde{z}$ is a preimage of z in $M(\alpha)$ and d is the differential of $M(\alpha)$, and we regard $d\tilde{z}$ as an element of the subcomplex G. But we may take $\tilde{z}$ to be $(z, 0) \in M(\alpha)_i = F_{i-1} \oplus G_i$, and then $d\tilde{z} = (0, \alpha_{i-1}(z))$, whence the assertion. □

Applications of the mapping cone to the proof of exactness of the Koszul complex and the Taylor complex are given in Chapter 17. A natural generalization of the mapping cone is the total complex of a double complex; we give the construction here, though we shall not use it seriously until we develop the language of spectral sequences.

For agreement with what we do later, we make a small change in notation. Up to this point we have usually dealt with complexes whose differentials d have degree -1:

$$\cdots \to F_n \xrightarrow{d} F_{n-1} \to \cdots .$$

In the interest of agreeing with most of the standard treatments of spectral sequences and double complexes, we shall now switch to complexes with differential of degree +1, and we shall write them with *upper* indices

$$\cdots \to F^m \to F^{m+1} \to \cdots .$$

If we take $m = -n$ and identify F_n with F^{-n}, we recover our previous notation. We shall generally adopt this convention for dealing with upper and lower indices. It has the advantage of avoiding negative indices. Thus we shall write an injective resolution of a module M as

$$0 \to M \to Q^0 \to Q^1 \to Q^2 \to \cdots,$$

and we can write a free resolution M either in the form

$$\cdots \to F^{-2} \to F^{-1} \to F^0 \to M \to 0$$

or in the old form

$$\cdots \to F_2 \to F_1 \to F_0 \to M \to 0.$$

I believe that the price of making such translations is more than repaid by the convenience, in dealing with spectral sequences, of always having the arrows point the same way.

Definition. *A **double complex** is a commutative diagram as in Figure A3.9 (extending infinitely in all four directions) where each row and each column is an ordinary complex; that is, a commutative diagram F as shown, with $d_{\text{hor}}^2 = 0 = d_{\text{vert}}^2$.*

Of course, any ordinary complex may be considered a double complex in which only one row is nonzero, and a map of ordinary complexes may be thought of as a double complex in which only two rows are nonzero. From the latter example, we have seen how to make an ordinary complex, the mapping cone. The natural generalization of this construction is a way of making an ordinary complex, called the associated **total complex**, from any double complex.

Definition. ***The total complex** of F is a complex whose kth term is*

$$\oplus_{i+j=k} F^{i,j},$$

with differential as in Figure A3.10. Somewhat more directly, one may think of a term of the total complex as the sum of the terms of the double complex along a diagonal, as shown by the line through the summands of $(\text{tot}\, F)^{i+j+1}$

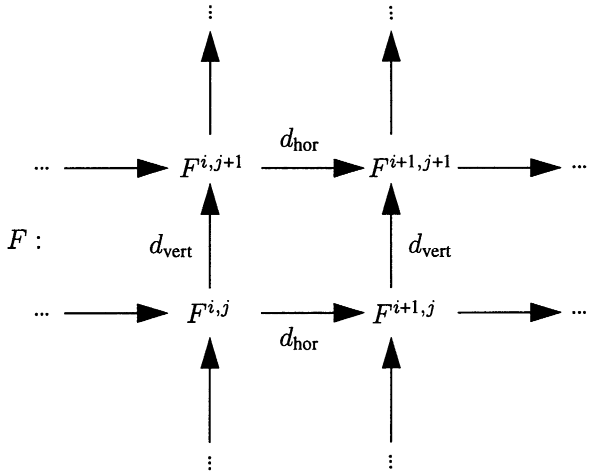


FIGURE A3.9.

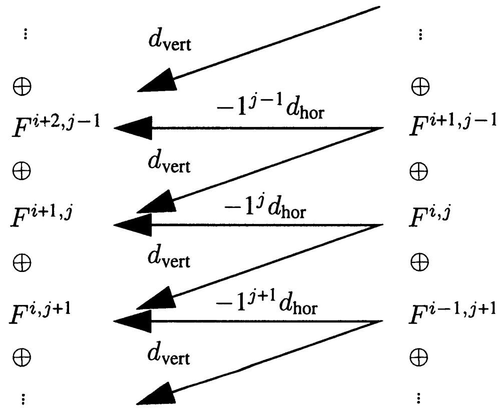


FIGURE A3.10.

in Figure A3.11. The differential is equal to the sum of all the maps shown, the maps in the jth row being multiplied by $(-1)^j$.

If F and G are ordinary complexes, with differentials φ and ψ, then the tensor product of F and G (as graded modules) becomes a double complex with terms $F_{i,j} = F^{-i,-j} := F_i \otimes G_j$ and differentials $d_{\text{hor}} = \varphi \otimes 1, d_{\text{vert}} = 1 \otimes \psi$, as in Figure A3.12.

Similarly, $\operatorname{Hom}(F, G)$ is a double complex with terms $F^{i,-j} := \operatorname{Hom}(F_i, G_j)$ and differentials $d_{\text{hor}} = \operatorname{Hom}(\varphi, 1)$ and $d_{\text{vert}} = \operatorname{Hom}(1, \psi)$. The homology of the total complex of $\operatorname{Hom}(F, G)$ has a nice interpretation, given in the following exercises.

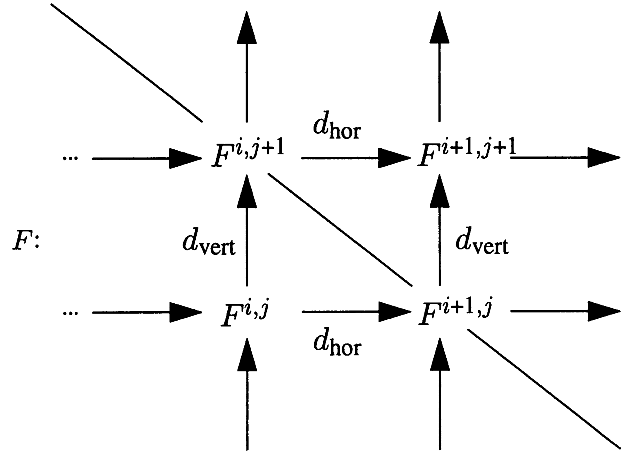


FIGURE A3.11.

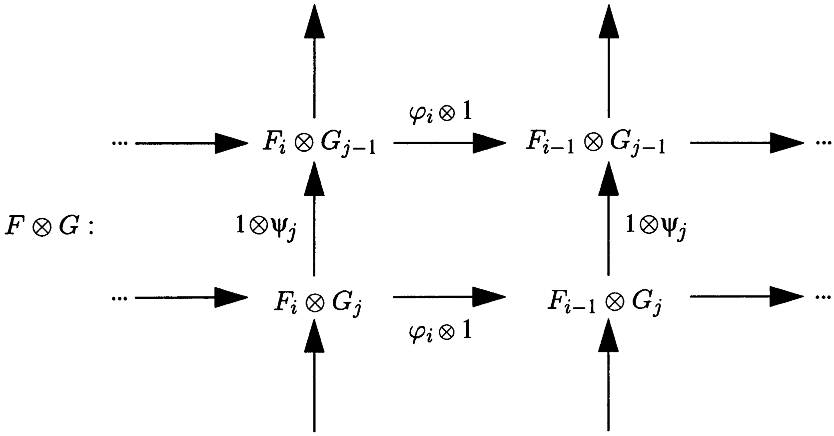


FIGURE A3.12.

A3.12.1 Exercises: Mapping Cones and Double Complexes

Exercise A3.30 (Resolution of an ideal from a factor ring): Suppose that R is a graded ring such that R_0 is a field, $I \subset R$ is an ideal, and J is an R-module. Suppose that

$$F: \cdots \to F_s \to \cdots \to F_1 \to R \to R/I \to 0$$

and

$$G: \cdots \to G_s \to \cdots \to G_1 \to G_0 \to J \to 0$$

are free resolutions of R/I and J. Given a monomorphism $a : J \to R/I$, identifying J with an ideal in R/I, let J' be the preimage of $a(J)$ in R. Given also maps $\alpha_i : G_i \to F_i$ forming a map of complexes $\alpha : F \to G$ lifting the map a, show that the mapping cone of α is a free resolution

$$M: \cdots \to F_s \oplus G_{s-1} \to \cdots \to F_1 \oplus G_0 \to R \to R/J' \to 0$$

of R/J'.

If R is local or graded, we would sometimes like to have a minimal free resolution of R/J'. Unfortunately, M need not be minimal even if F and G are, but there is one moderately common case in which we can prove that M is minimal. Suppose that $R = R_0 \oplus R_1 \oplus \cdots$ is a graded ring, and write each F_i and G_j as a sum of twists of $R: F_i = \oplus_j R(f_{ij})$ and $G_i = \oplus_j R(g_{ij})$. Show that if $f_{ij} > g_{ik}$ for all i, j, k, then M is minimal.

Exercise A3.31: If α is an isomorphism of complexes, show that the complex $M(\alpha)$ is "homotopically trivial" in the sense that the identity map from $M(\alpha)$ to itself is homotopic to the zero map.

Exercise A3.32: A **quasi-isomorphism of complexes** is a map of complexes that induces an isomorphism on homology; two complexes are quasi-isomorphic if there is a quasi-isomorphism between them (in either direction). A homotopy equivalence of complexes F and G is a map $\alpha : F \to G$ such that there is a map $\beta : G \to F$ with the property that $\alpha\beta$ and $\beta\alpha$ are each homotopic to the identity. Show that a homotopy equivalence is a quasi-isomorphism. Show by example that not every quasi-isomorphism is a homotopy equivalence. Show by example that two complexes may have the same homology without being quasi-isomorphic.

Exercise A3.33: Suppose that

$$0 \to F' \xrightarrow{\alpha} F \xrightarrow{\beta} F'' \to 0$$

is a short exact sequence of complexes. Show that F'' is quasi-isomorphic to $M(\alpha)$ by showing that there is a short exact sequence of complexes

$$0 \to M(\alpha') \to M(\alpha) \to F'' \to 0,$$

where α' is the isomorphism of F' onto $\alpha(F') \subset F$, and using Exercise A3.32. Similarly, show that F' is quasi-isomorphic to $M(\beta)$ (up to a shift of degree).

Exercise A3.34: Show that if

$$F: \cdots \to F^{n-2} \to F^{n-1} \to F^n \to 0 \to 0 \to \cdots$$

is a complex "bounded above" and

$$G: \cdots \to 0 \to 0 \to G^m \to G^{m+1} \to G^{m+2} \to \cdots$$

is a complex "bounded below," then the cycles of degree i in $\mathrm{tot}(\mathrm{Hom}(F, G))$ are the degree-i maps of complexes from F to G (that is, collections of maps $F_{-j} = F^j \to G^{j+1}$ that commute with the differentials), and the boundaries are the maps homotopic to 0; thus $H_i(\mathrm{tot}(\mathrm{Hom}(F, G)))$ is the group of homotopy classes of maps of degree i from F to G. The same thing is true if F is bounded below and G is bounded above.

A3.13 Spectral Sequences

General references for spectral sequences:[2] Serre [1957] does the case of a filtered complex quite directly (I learned the subject from this source). Other good treatments may be found in Maclane [1963], Cartan and Eilenberg [1956], Godement [1958], Grothendieck [1957] and Hilton and Stammbach [1971]. For a particularly gentle exposition of the subject with topological intentions, see Bott and Tu [1982] (but watch out for misprints).

Spectral sequences first arose in the work of Leray [1946, 1950] on topology and independently in the work of Lyndon [1946, 1948] on group cohomology. The topologists are the primary consumers of the theory, but there are plenty of applications in commutative algebra, in various algebraic cohomology theories, and in other areas as well.

It is easy to describe a spectral sequence.

Definition. *A* ***spectral sequence*** *is a sequence of modules* ${}^rE$ *for* $r \geq 1$, *each with a "differential"* $d_r : {}^rE \to {}^rE$ *satisfying* $d_r d_r = 0$, *such that* ${}^{r+1}E \cong \ker d_r / \operatorname{im} d_r$ *(or, as we shall prefer to write it,* ${}^{r+1}E = H({}^rE)$, *the homology of* ${}^rE$*).*

From these data one can define a "limit" term ${}^\infty E$. A spectral sequence may be interesting because ${}^\infty E$ may be identified with some inherently interesting object, to which the ${}^rE$ become "successive approximations"; or, on occasion, because the ${}^rE$ are interesting and ${}^\infty E$ is somehow trivial, which shows that some of the maps d_r must be very nontrivial.

To define ${}^\infty E$, we first define submodules

$$0 = {}^1B \subset {}^2B \subset \cdots \subset {}^rB \subset \cdots \subset \cdots \subset {}^rZ \subset \cdots \subset {}^2Z \subset {}^1Z = {}^1E$$

such that ${}^iE = {}^iZ/{}^iB$ for each i. To do this, let ${}^1Z = {}^1E$, and ${}^1B = 0$, so that ${}^1E = {}^1Z/{}^1B$. Having defined ${}^iB$ and ${}^iZ$, for $i \leq r$ we define ${}^{r+1}Z$ as the kernel of the composite map

$${}^rZ \to {}^rZ/{}^rB = {}^rE \xrightarrow{d_r} {}^rE = {}^rZ/{}^rB,$$

and write the image of this map as ${}^{r+1}B/{}^rB$; clearly ${}^{r+1}Z/{}^{r+1}B = H({}^rE) = {}^{r+1}E$, and

$${}^rB \subset {}^{r+1}B \subset {}^{r+1}Z \subset {}^rZ,$$

as required. Having defined all the ${}^iZ$ and ${}^iB$, we set

$${}^\infty Z = \bigcap^r Z,$$

$${}^\infty B = \bigcup^r B;$$

[2]Spectral sequences = suites spectrales; and spectral sweets = ghost candy.

and finally we define the **limit** of the spectral sequence to be

$$^{\infty}E = {^{\infty}Z}/{^{\infty}B}.$$

We say that the spectral sequence **collapses at** $^{r}\boldsymbol{E}$ if $^{r}E = {^{\infty}E}$, or equivalently if the differentials $d_r, d_{r+1}, d_{r+2}, \ldots$ are 0.

Where do interesting spectral sequences come from? Most of the applications in algebra have to do with a spectral sequence that arises from a double complex in a way to be described shortly, a construction that generalizes the theory of the mapping cone that we have already used. There are also a few applications of the more general notion of the spectral sequence of a filtered complex. Still more general is a construction introduced by Massey [1952] that derives a spectral sequence from an object called an exact couple. There is an exact couple associated to any monomorphism from one complex (or differential module) to another, and it seems that most useful spectral sequences can be defined this way.

The subject of spectral sequences is elementary, but the notion of the spectral sequence of a double complex involves so many objects and indices that it seems at first repulsive. The approach via exact couples allows a much simpler view, postponing the indices until they are really needed; we shall follow this approach. First, however, we introduce the subject by recasting the theory of the mapping cone in the form it takes as a special case of the theory of the spectral sequence of a double complex.

A3.13.1 Mapping Cones Revisited

Suppose that $\alpha : F \to G$ is a map of complexes, and that we are interested in the homology of the mapping cone $M := M(\alpha)$. We shall show that the long exact sequence in homology of Proposition A3.19 can be interpreted as giving a filtration on the homology of M and a (very simple) spectral sequence whose $^{\infty}E$ term is the associated graded module of this filtration. This is a special case of the situation that holds more generally for (reasonable) double complexes.

The complex M contains a subcomplex M^1 isomorphic to G, with quotient $M/M^1 \cong F[-1]$. The resulting long exact sequence in homology has the form

$$\cdots \to H_iF \xrightarrow{\alpha_i} H_iG \to H_iM \to H_{i-1}F \xrightarrow{\alpha_{i-1}} H_{i-1}G \to \cdots,$$

where we have written α_i and α_{i-1} for the maps induced on homology. Saying that there is such an exact sequence is equivalent to saying that H_iM has a filtration, which we shall write as

$$H_iM = (H_iM)^0 \supset (H_iM)^1 \supset (H_iM)^2 = 0,$$

where $(H_iM)^1 = \operatorname{im} H_iG \to H_iM$, such that

$$(H_iM)^1/(H_iM)^2 = \operatorname{coker} \alpha_i,$$
$$(H_iM)^0/(H_iM)^1 = \ker \alpha_{i-1}.$$

We write $HF = \oplus_i H_iF$, and similarly for G and M. Write α_* for the direct sum of the maps α_i, so that $\alpha_* : HF \to HG$.

We can now define the spectral sequence: Let ${}^1E$ be the module $HF[-1] \oplus HG$. The module ${}^1E$ has a "differential" d_1 that is the composite

$${}^1E : HF[-1] \oplus HG \twoheadrightarrow HF[-1] \xrightarrow{\alpha_*} HG \rightarrowtail HF[-1] \oplus HG,$$

where the left-hand map is projection onto the first factor, and the right-hand map is injection into the second factor. It is clear that $\ker d_1 = \ker \alpha_* \oplus HG$, and $\operatorname{im} d_1 = 0 \oplus \operatorname{im} \alpha_*$, so

$${}^2E := H({}^1E) = \ker \alpha_* \oplus \operatorname{coker} \alpha_*.$$

We give ${}^2E$ and all the succeeding ${}^rE$ the differential 0, so that the resulting spectral sequence collapses at ${}^2E$, and ${}^2E = {}^3E = \cdots = {}^\infty E$. The above relations may thus be written as

$$\operatorname{gr} HM = H^\infty E,$$

where gr HM is the associated graded module of HM, that is,

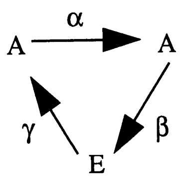
$$\operatorname{gr} HM := (HM)^0/(HM)^1 \oplus (HM)^1/(HM)^2.$$

This is the form that is generalized to arbitrary double complexes and beyond in the next section.

A3.13.2 Exact Couples

An **exact couple** is an **exact triangle**[3] of the form

(*)



[3] The reader who objects to defining an exact couple to be an exact triangle has my sympathy. Presumably the fact that there are only two distinct modules in the triangle, A and E, is the origin of the name.

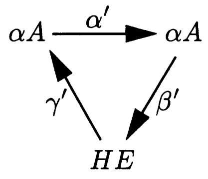
—that is, a diagram of modules and maps as above, which is exact in the obvious sense that $\ker\alpha = \operatorname{im}\gamma, \ker\gamma = \operatorname{im}\beta$, and $\ker\beta = \operatorname{im}\alpha$. Let $d : E \to E$ be the composite map $d = \beta\gamma$. Since $\gamma\beta = 0$, we see that $d^2 = 0$, so E is a differential module, and we write

$$HE = \ker d / \operatorname{im} d$$

for its homology.

Proposition–Definition A3.20. *If the diagram* $(*)$ *above is an exact couple, then there is a* ***derived exact couple***

(**)



where:

α' *is* α *restricted to* αA*, the image of* α*;*

β' *is* $\beta \circ \alpha^{-1} : \alpha A \to HE$*, taking* αa *to the homology class of* βa*;*

γ' *is the map induced by* γ *on* $\ker d$ *(which automatically kills* $\operatorname{im} d$*).*

Proof. Note that β' is well defined because $\ker\alpha = \operatorname{im}\gamma$ is taken to $\operatorname{im} d$ by β. The proof of exactness is completely straightforward, and we leave it to the reader. □

Given an exact couple $(*)$ we may form the derived exact couple $(**)$, and then repeat the process on $(**)$ … . Thus we get the **spectral sequence of the exact couple,** defined by:

$^1E = E$ with differential $d_1 = d = \beta\gamma$, from the original couple;
$^2E = HE$ with differential $d_2 = \beta'\gamma'$, from the derived couple;
$^3E = HHE \cdots$ from the derived couple of the derived couple;

and so forth.

It is easy to check that with notation as in the definition of a spectral sequence we have

$$^{r+1}Z = \gamma^{-1}(\operatorname{im}\alpha^r)$$
$$^{r+1}B = \beta(\ker\alpha^r),$$

where α^r is the composite of α with itself r times. Thus

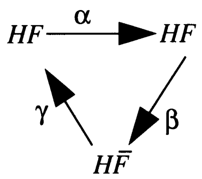
$$
\begin{aligned}
{}^{\infty}E &= {}^{\infty}Z/{}^{\infty}B \\
&= \frac{\gamma^{-1}\left(\bigcap_r \operatorname{im}\alpha^r\right)}{\beta\left(\bigcup_r \ker\alpha^r\right)}.
\end{aligned}
$$

Where do interesting exact couples come from? All of those treated here are instances of the following construction:

Let F be a differential module over a ring R, and let $\alpha : F \to F$ be a monomorphism. Set $\bar{F} = F/\alpha F$. The module $\bar{F}$ inherits a differential from F, so the short exact sequence of differential modules

$$0 \to F \xrightarrow{\alpha} F \to \bar{F} \to 0$$

gives rise to an exact triangle in homology



where we have written α again for the map on homology induced by $\alpha : F \to F$. The spectral sequence of this exact couple will be called the **spectral sequence of α on F**.

It is convenient to think of the map α as induced by multiplication with an element α of R that is a nonzerodivisor on F. Every case may be regarded this way—if necessary we adjoin a new variable α to R, and let it act as α on F (and thus also on HF), so that F and $\bar{F}$ become $R[\alpha]$-modules, with $\bar{F} = F/\alpha F$. If R is $\mathbf{Z}$, the ring of integers, and $\alpha \in \mathbf{Z}$ is an integer, then the spectral sequence above is widely known as the Bockstein spectral sequence, and the differentials as the Bockstein operators, but much of the theory is the same in the general case. With this in mind, we shall call $\ker \alpha^r : HF \to HF$ the **α^r-torsion of HF.** We shall also consider the intermediate complexes $F/\alpha^r F$; we say that a class in $H\bar{F}$ can be **lifted modulo α^r** if it is in the image of the natural map $H(F/\alpha^r F) \to H\bar{F}$; that is, if it has a representative in F (not necessarily a cycle) that becomes a cycle modulo α^r.

Proposition A3.21. *In the spectral sequence of α on F, the module ${}^{r+1}Z$ is the set of classes in $H\bar{F}$ that can be lifted modulo α^{r+1}, while ${}^{r+1}B$ is the image in $H\bar{F}$ of the α^r-torsion in HF. If $\bar{z}$ is a cycle in $\bar{F}$ with a representative $z \in F$ that is a cycle modulo α^r, then the differential $d_{r+1} : {}^{r+1}E \to {}^{r+1}E$ takes the class $\bar{z}$ to the class of $\alpha^{-(r+1)}dz$, where d is the differential of F.*

Proof. If z is any lifting to F of a cycle $\bar{z}$ in $H\bar{F}$, then $\gamma[\bar{z}] = [\alpha^{-1}dz] \in HF$. Further if $z \in F$ represents a cycle in $F/\alpha^r F$, then dz is divisible by α^r, so $\alpha^{-r}dz \in F$ makes sense; it is a cycle because α is a monomorphism on F. The rest is immediate from the definitions. □

A3.13.3 Filtered Differential Modules and Complexes

A **filtered differential module** is a differential module (G, d) together with a sequence of submodules G^p satisfying

$$G \supset \cdots \supset G^p \supset G^{p+1} \supset \cdots, \qquad p \in \mathbf{Z}$$

that are preserved by d —that is, $dG^p \subset G^p$ for all p. If in addition G is graded (for example, G might be a complex), say by upper degrees $G = \oplus_q G^q$, then we write $(G^q)^p$ for the pth level in the filtration of G^q. There are two examples that the reader should bear in mind. Recall that we write G_q for G^{-q}.

Example A. Let

$$G : \cdots \to G_{q+1} \xrightarrow{\varphi_{q+1}} G_q \xrightarrow{\varphi_q} G_{q-1} \to \cdots$$

be a complex of finitely generated modules over a Noetherian local ring $(R, \mathfrak{m})$, and let

$$G^p = \mathfrak{m}^p G : \cdots \to \mathfrak{m}^p G_{q+1} \xrightarrow{\varphi_{q+1}} \mathfrak{m}^p G_q \xrightarrow{\varphi_q} \mathfrak{m}^p G_{q-1} \to \cdots .$$

For the interesting applications we shall need more general filtrations

$$G^p : \cdots \to G^p_{q+1} \xrightarrow{\varphi_{q+1}} G^p_q \xrightarrow{\varphi_q} G^p_{q-1} \to \cdots,$$

satisfying only the property that $\cdots \supset G^p_q \supset G^{p+1}_q \supset \cdots$ is an $\mathfrak{m}$-stable filtration in the sense of Chapter 5 and Exercise A3.42.

We regard G as a filtered differential module by taking the direct sum over all q, as usual.

Example B. Let F be a double complex as in Figure A3.9 and let G be the total complex, $G = \operatorname{tot} F$. There are two natural filtrations on G—vertical filtration and horizontal filtration. The horizontal filtration is defined by subcomplexes ${}_{\text{hor}}G^p$, where ${}_{\text{hor}}G^p$ comes from the rows of F where the second index $\geq p$; that is ${}_{\text{hor}}G^p$ is made from the rows from $F^{*,p}$ up, shaded in Figure A3.13.

Similarly, ${}_{\text{vert}}G^p$ is the subcomplex coming from the columns where the first index is $\geq p$; in the picture, these are the columns $F^{p,*}$ and to the right. More formally, we let

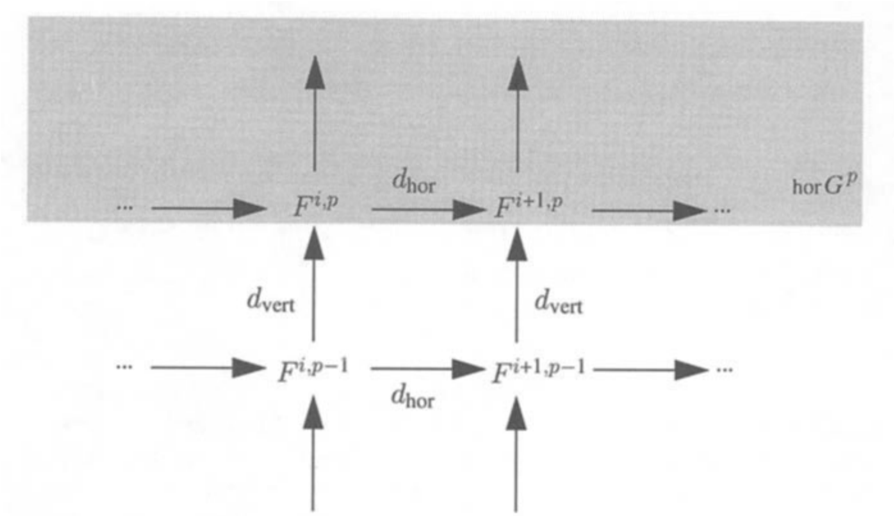


FIGURE A3.13.

$$\left({}_{\mathrm{hor}}G^p\right)^k = \oplus_{i+j=k,\ j\geq p} F^{i,j}$$

with differential defined as the restriction of the differential of G, and similarly for ${}_{\mathrm{vert}}G^p$.

In this section we shall give a general procedure for making a spectral sequence from a filtered differential module and we shall consider Example A. In the next section we shall consider Example B. In each of these cases we simply interpret Proposition A3.21; it is a new interpretation of the limit term that makes these situations interesting.

Let G be a filtered differential module as above, and let $F = \oplus_{p\in\mathbf{Z}} G^p$. The sum of the inclusion maps $G^{p+1} \to G^p$ defines a map $\alpha : F \to F$ that is obviously a monomorphism. Its cokernel $\bar{F}$ is obviously

$$\mathrm{gr}\, G := \oplus_p G^p/G^{p+1}.$$

Thus, setting ${}^1E^p = H(G^p/G^{p+1})$, we see that the spectral sequence of α on F starts with

$${}^1E = H(\mathrm{gr}\, G) = \oplus_p H(G^p/G^{p+1}) = \oplus_p {}^1E^p.$$

Since $F/\alpha^r F = \oplus_p G^p/G^{p+r}$, we may interpret Proposition A3.21 as saying in this case that ${}^{r+1}Z = \oplus_p {}^{r+1}Z^p$, with

$$\begin{aligned} {}^{r+1}Z^p &= \{[z] \in {}^1E^p | z \in G^p \text{ and } dz \in G^{p+r+1}\} \\ &= \{z \in G^p | dz \in G^{p+r+1}\} + G^{p+1}/G^{p+1} + dG^p; \\ {}^{r+1}B^p &= \{[z] \in {}^1E^p | z \in G^p \text{ and } z = dy \quad \text{for some } y \in G^{p-r}\} \\ &= (G^p \cap dG^{p-r}) + G^{p+1}/G^{p+1} + dG^p. \end{aligned}$$

So far, this is nothing but an application of Proposition A3.21. The new element is the following relation of ${}^{\infty}E$ with HG. The module HG is filtered by the submodules $(HG)^p = \operatorname{im} H(G^p) \to HG$. The associated graded module may be written as gr $HG = \oplus_p (HG)^p/(HG)^{p+1}$, and writing $K^p = \{z \in G^p | dz = 0\}$, we have

$$\begin{aligned}(HG)^p/(HG)^{p+1} &= K^p/(K^{p+1} + (dG \cap G^p)) \\ &= K^p + G^{p+1}/(G^p \cap dG) + G^{p+1}\end{aligned}$$

because $K^p \cap ((G^p \cap dG) + G^{p+1}) = K^{p+1} + (dG \cap G^p)$.

This last expression for $(HG)^p/(HG)^{p+1}$ is quite similar to the expression

$${}^{\infty}Z^p/{}^{\infty}B^p = \bigcap_r (\{z \in G^p | dz \in G^{p+r}\} + G^{p+1}) / \bigcup_r ((G^p \cap dG^{p-r}) + G^{p+1}).$$

Writing the quotient on the right as M^p/N^p, we have

$$\begin{aligned}M^p &\supset K^p + G^{p+1}, \\ N^p &\subset (G^p \cap dG) + G^{p+1};\end{aligned}$$

so taking the direct sum over all p, we get

$$\operatorname{gr} HG \text{ is a quotient of a submodule of } {}^{\infty}Z/{}^{\infty}B.$$

Definition. *We say that the spectral sequence of the filtered differential module G* ***converges****, and for any term ${}^{r}E$ of the spectral sequence we write* ${}^{r}\boldsymbol{E} \Rightarrow$ **gr** $\boldsymbol{HG}$, *if gr $HG = {}^{\infty}Z/{}^{\infty}B$; that is, if for each p we have*

i. $\cap_r(\{z \in G^p | dz \in G^{p+r}\} + G^{p+1}) = \{z \in G^p | dz = 0\} + G^{p+1}$, *and*

ii. $\cup_r((G^p \cap dG^{p-r}) + G^{p+1}) = (G^p \cap dG) + G^{p+1}$.

Note that condition ii is relatively trivial; it will be satisfied as soon as $G = \cup_p G^p$. Condition i, however, is much more subtle.

Theorem A3.22. *The spectral sequence in Example A converges; further, the filtration induced on the homology of G is $\mathfrak{m}$-stable.*

Proof. We prove convergence for the $\mathfrak{m}$-adic filtration, leaving the important generalization to the reader in Exercise A3.42. In this spectral sequence, $G^p = G$ for $p \le 0$; thus convergence condition ii is trivially satisfied.

We now turn to condition i, which we may rewrite in the form

$$\begin{aligned}\bigcap_r (\{z \in G^p | dz \in G^{p+r}\} + G^{p+1})/\{z \in G^p | dz = 0\} \\ = (\{z \in G^p | dz = 0\} + G^{p+1})/\{z \in G^p | dz = 0\}.\end{aligned}$$

The proof uses the Artin-Rees lemma (Lemma 5.1) and the Krull intersection theorem (Corollary 5.4).

Note that each G^p is a direct sum of the finitely generated modules $(G_q)^p = \mathfrak{m}^p G_q$, and the result we want may be checked for one of these summands at a time.

First, set $X = (G_q)^p/\{z \in (G_q)^p | dz = 0\}$. The differential d induces an inclusion $X \subset (G_{q-1})^p$, and for sufficiently large r',

$$\begin{aligned}\{z \in (G_q)^p \ | dz \in (G_{q-1})^{p+r'+r}\}/\{z \in G^p | dz = 0\} \\ \subset X \cap (G_{q-1})^{p+r} \\ = X \cap \mathfrak{m}^r (G_{q-1})^p.\end{aligned}$$

By the Artin-Rees lemma, there is a number s such that this is contained in $\mathfrak{m}^{r-s}X$ for all $r \geq s$, so that

$$\{z \in (G_q)^p | dz \in (G_{q-1})^{p+r'+r}\} \subset \mathfrak{m}^{r-s}(G_q)^p + \{z \in (G_q)^p | dz = 0\}.$$

Thus

$$\begin{aligned}&\{z \in (G_q)^p | dz \in (G_{q-1})^{p+r+r'}\} + (G_q)^{p+1}/\{z \in (G_q)^p | dz = 0\} \\ &\subset \mathfrak{m}^{r-s}(G_q)^p + \{z \in (G_q)^p | dz = 0\} + (G_q)^{p+1}/\{z \in (G_q)^p | dz = 0\},\end{aligned}$$

and the intersection of these for all r is

$$\{z \in G_u^p | dz = 0\} + G_u^{p+1}/\{z \in G_u^p | dz = 0\}$$

by the Krull intersection theorem. We leave the $\mathfrak{m}$-stability of the induced filtration on HG to the reader (see Exercise A3.42).

The following corollary contains two simple applications.

Corollary A3.23.

a. *Let $x_1, \ldots, x_r$ be a sequence of elements in the maximal ideal $\mathfrak{m}$ of a local ring $(R, \mathfrak{m})$, and write x_i^* for the leading form of x_i in $\operatorname{gr}_{\mathfrak{m}} R$. If the x_i^* form a regular sequence on $\operatorname{gr}_{\mathfrak{m}} R$, then the x_i form a regular sequence on R.*

b. *Let M and N be finitely generated modules over a local ring $(R, \mathfrak{m})$. There is a spectral sequence*

$$\operatorname{Tor}^{\operatorname{gr} R}(\operatorname{gr} M, \operatorname{gr} N) \Rightarrow \operatorname{Tor}^R(M, N).$$

Thus, for example, if $\operatorname{Tor}_u^{\operatorname{gr} R}(\operatorname{gr} M, \operatorname{gr} N) = 0$ then $\operatorname{Tor}_u^r(M, N) = 0$.

Proof. Follow the hints in Exercises A3.43 and A3.44. □

We emphasize that the filtration induced on $\operatorname{Tor}^R(M, N)$ in this corollary may not be the $\mathfrak{m}$-adic filtration, but will be $\mathfrak{m}$-stable.

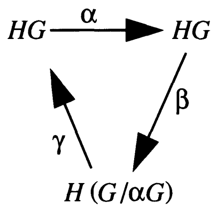
A3.13.4 The Spectral Sequence of a Double Complex

We now take up Example B, which is arguably the most important for algebraists. For this discussion we keep the example's notation, summarized in Figure A3.9, with $G = \operatorname{tot} F$.

We consider HG and $H(G/\alpha G)$ as bigraded modules by setting

$$(HG)^{p,q} := H^{p+q}(G^p) \text{ and } H(G/\alpha G)^{p,q} := H^{p+q}(G^p/G^{p+1}).$$

The maps α, β, and γ in the exact triangle



are then bigraded of degrees $(-1,1)$, $(0,0)$, and $(1,0)$, respectively. Thus the differential $d_r : {}^rE \to {}^rE$ is bihomogeneous of degree r in the p grading and $-(r-1)$ in the q grading; that is, d_r is the direct sum of maps

$$d_r : {}^rE^{p,q} \to {}^rE^{p+r,q-r+1}.$$

More graphically, representing ${}^1E$ as an array of the ${}^1E^{p,q}$,

${}^1E^{p,q}$	${}^1E^{p+1,q}$	${}^1E^{p+2,q}$
${}^1E^{p,q-1}$	${}^1E^{p+1,q-1}$	${}^1E^{p+2,q-1}$
${}^1E^{p,q-2}$	${}^1E^{p+1,q-2}$	${}^1E^{p+2,q-2}$

the differential d_r goes "r steps to the right and $r-1$ steps down" as in Figure A3.14.

Of course, this picture needs some interpretation: d_2 is actually defined on the kernel of d_1 (a quotient of which is ${}^2E^{*,*}$); d_3 is actually defined on the kernel of d_2; and so on. To describe the d_i, suppose for definiteness that we are working with the spectral sequence of the horizontal filtration. Then d_1 is simply the map induced by d_{hor} on the homology of d_{vert}. An element of the kernel of d_1 is represented by a "vertical cycle" $z \in F^{p,q}$ (that is, an element of the kernel of d_{vert}) that is mapped by d_{hor} to 0 in homology —that is, such that $d_{\text{hor}}(z)$ is a "vertical boundary," an element of the form $d_{\text{vert}}(z')$. The map d_2 takes (the homology class of) z to the homology class of $d_{\text{hor}}(z')$. For this to be zero means that $d_{\text{hor}}(z') = d_{\text{vert}}(z'')$ for some z'', and in this case d_3 carries (the homology class of) z to $d_{\text{hor}}(z'')$; and so on.

Here is our main result.

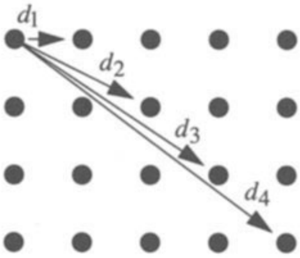


FIGURE A3.14.

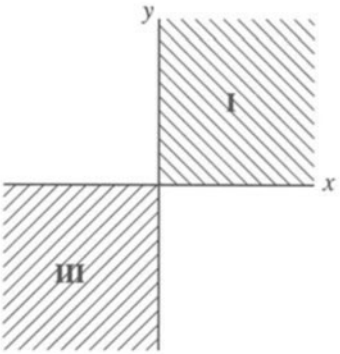


FIGURE A3.15.

Theorem A3.24. *Associated with the double complex F are two spectral sequences, ${}^r_{\text{hor}}E$ and ${}^r_{\text{vert}}E$, corresponding, respectively, to the horizontal and vertical filtrations of* $\operatorname{tot}(F) = G$. *The ${}^1E$ terms are bigraded with the components given by*

$$ {}^1_{\text{hor}}E^{p,q} = H^q(F^{*,p}), \qquad {}^1_{\text{vert}}E^{p,q} = H^q(F^{p,*}). $$

If $F^{i,j} = 0$ for all $i < 0$ or for all $j > 0$, then the horizontal spectral sequence converges; that is,

$$ {}^\infty_{\text{hor}}E = \operatorname{gr}_{\text{hor}} H(\operatorname{tot} F). $$

Symmetrically, if $F^{i,j} = 0$ for all $i > 0$ or for all $j < 0$, then the vertical spectral sequence converges.

Terminology: Theorem A3.24 implies that both spectral sequences converge either if $F^{i,j} = 0$ for all $i < 0$ and for all $j < 0$ or if $F^{i,j} = 0$ for all $i > 0$ and for all $j > 0$. In the former case, the nonzero terms are all in the first quadrant of the i, j-plane, and we call F a **first-quadrant double complex**. In the second case the nonzero terms are all in the third quadrant, and we call F a **third-quadrant double complex** (see Figure A3.15).

Proof. The proof of the first formula is immediate from the definitions. For example, we have

$$\begin{aligned} {}_{\text{hor}}^{\;1}E^{p,q} &= H^{p+q}(\mathrm{gr}_{\text{hor}}(\operatorname{tot} F)^p) \\ &= H^q(F^{*,p}), \end{aligned}$$

whence the formula for ${}_{\text{hor}}^{\;1}E^{p,q}$; the case of the vertical spectral sequence is similar.

The proof of convergence uses the bigrading. Writing G for the total complex $\operatorname{tot} F$, we must show that

i. $\cap_r(\{z \in G^p | dz \in G^{p+r}\} + G^{p+1}) = \{z \in G^p | dz = 0\} + G^{p+1}$

and

ii. $\cup_r((G^p \cap dG^{p-r}) + G^{p+1}) = (G^p \cap dG) + G^{p+1}$.

Since $G = \cup_p G^p$ with respect to either filtration, condition ii is trivially satisfied (this does not use any conditions on the double complex). Condition i means that if $z \in G^p$ and for each r there is an element $y_r \in G^{p+1}$ such that $d(z - y_r) \equiv 0 \bmod G^{p+r}$, then there is some $y \in G^{p+1}$ such that $d(z - y) = 0$. In our case, since G is a complex, it is enough to check this for $z \in (G^q)^p$, for some q. For definiteness, consider again the case of the horizontal filtration. The element $d(z - y_r)$ is then in

$$(G^{q+1})^{p+r} = \bigoplus_{i+j=q+1, j\geq p+r} F^{i,j}.$$

If $F^{i,j} = 0$ for $j > 0$, then $(G^*)^{p+r} = 0$ for $r > -p$. If, on the other hand, $F^{i,j} = 0$ for $i < 0$, then $(G^{q+1})^{p+r} = 0$ if $r > q + 1 - p$. Thus in either case $d(z - y_r) = 0$ for suitable r, and we may take $y = y_r$ for this value of r.

A refinement of the notation for convergence is useful: We write

$${}^rE^{p,q} \Rightarrow_p H^{p+q}(\operatorname{tot} F)$$

to mean that the spectral sequence containing the terms ${}^rE^{p,q}$ converges, and that writing $H^{p+q}(\operatorname{tot} F)^p$ for the pth level in the associated filtration of $H^{p+q}(\operatorname{tot} F)$,

$${}^\infty E^{p,q} = H^{p+q}(\operatorname{tot} F)^p / H^{p+q}(\operatorname{tot} F)^{p+1}.$$

We now give two simple applications. More will be found in the exercises. The first involves a double complex both of whose spectral sequences degenerate at ${}^2E$.

i. Balanced Tor. We shall show that $\operatorname{Tor}_i^R(M, N)$ may be computed from a free resolution of either M or N and is in fact a "balanced" functor in the

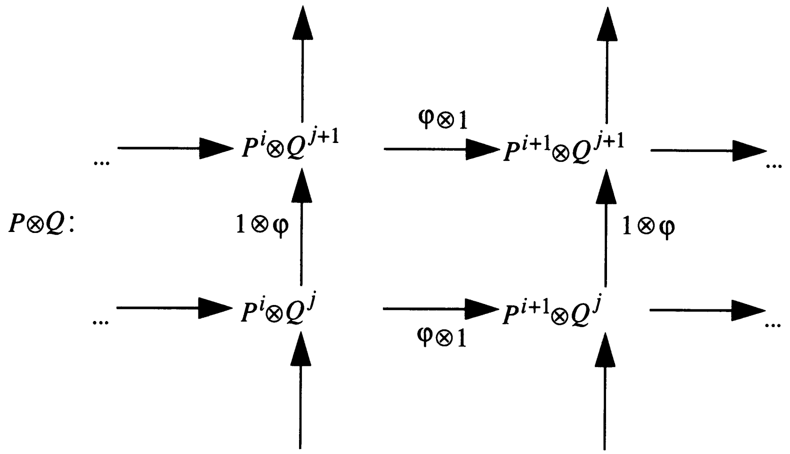


FIGURE A3.16.

sense that if $a \in R$, then multiplication by a on M induces the same map on $\operatorname{Tor}_i^R(M,N)$ as does multiplication by a on N—that is, the R-module structure on $\operatorname{Tor}_i^R(M,N)$ may be induced from the module structure of either M or N. To see this let

$$P: \cdots \to P_i \xrightarrow{\varphi_i} P_{i-1} \to \cdots \to P_0$$

and

$$Q: \cdots \to Q_i \xrightarrow{\psi_i} Q_{i-1} \to \cdots \to Q_0$$

be free resolutions of M and N, respectively. We shall show that

$$H(P \otimes_R N) \cong H(\operatorname{tot}(P \otimes_R Q)) \cong H(M \otimes_R Q),$$

as R-modules. Since "$\operatorname{Tor}_i^R(M,N)$ computed from a free resolution of M" is the first of these, and "$\operatorname{Tor}_i^R(M,N)$ computed from a free resolution of N" is the last, this will suffice.

Let ${}_{\text{vert}}E$ be the vertical spectral sequence associated with the third-quadrant double complex $F = P \otimes_R Q$, which may be written with upper indices, using the convention that $P^i = P_{-i}$, in the form shown in Figure A3.16. We have ${}_{\text{vert}}^{\ 1}E_{i,j} = {}_{\text{vert}}^{\ 1}E^{-i,-j} = H_j(P_i \otimes Q)$. Since P_i is free, the complex $P_i \otimes Q$ is just a direct sum of copies of Q; more invariantly, we have $H_j(P_i \otimes Q) = P_i \otimes H_j(Q)$. This is 0 for $j > 0$, while $P_i \otimes H_0(Q) = P_i \otimes N$. Thus the only nonzero ${}_{\text{vert}}^{\ 1}E_{i,j}$ are those with $j = 0$. The differential d_1 is induced by $d_{\text{hor}} = \varphi \otimes 1$. Thus ${}_{\text{vert}}^{\ 1}E$ is the complex $P \otimes N$, and

$${}_{\text{vert}}^{\ 2}E_{i,j} = \begin{cases} H_i(P \otimes N) & \text{for } j = 0 \\ 0 & \text{for } j > 0. \end{cases}$$

It follows that the spectral sequence degenerates at ${}^2E$; that is, ${}_{\text{vert}}^{\infty}E = {}_{\text{vert}}^{2}E$. Since all the nonzero terms have $j = 0$,

$$\bigoplus_{i+j=k} {}_{\text{vert}}^{\infty}E_{i,j} = {}^{\infty}E_{k,0},$$

and the filtration of $H(\text{tot}(P \otimes Q))$ has only one nonzero piece. Thus we get $H(P \otimes N) = H(\text{tot}(P \otimes Q))$. By symmetry, $H(\text{tot}(P \otimes Q)) = H(M \otimes Q)$ as well; we could also deduce this from the horizontal spectral sequence of $P \otimes Q$.

ii. Change of Rings. Let $R \to S$ be a ring homomorphism, let A be an S-module, and let B be an R-module. We shall derive one of the "change of rings spectral sequences" (see Exercise A3.45 for others), whose ${}^2E^{i,j}$ term is $\text{Ext}_S^i(A, \text{Ext}_R^j(S, B))$, converging to $\text{Ext}_R^{i+j}(A, B)$; that is,

$$\text{Ext}_S^i(A, \text{Ext}_R^j(S, B)) \Rightarrow_i \text{Ext}_R^{i+j}(A, B).$$

Let

$$P: \ \cdots \to P_i \to P_{i-1} \to \cdots \to P_0$$

be an S-free resolution of A as an S-module, and let

$$Q: Q^0 \to \cdots \to Q^j \to Q^{j+1} \to \cdots$$

be an R-injective resolution of B as an R-module, respectively. We regard $\text{Hom}_S(P, \text{Hom}_R(S, Q))$ as a first-quadrant double complex, with $\text{Hom}_S(P_i, \text{Hom}_R(S, Q^j))$ as the i, j term. We first claim that the horizontal spectral sequence degenerates, as in Example i. We have

$${}_{\text{hor}}^{1}E^{j,i} = H^i(\text{Hom}_S(P_*, \text{Hom}_R(S, Q^j)).$$

Since Q^j is R-injective, $\text{Hom}_R(S, Q^j)$ is S-injective, and

$$H^i(\text{Hom}_S(P_*, \text{Hom}_R(S, Q^j)) = \text{Hom}_S(H_i(P_*), \text{Hom}_R(S, Q^j)).$$

Since P_* is a resolution of A, this vanishes except for $i = 0$, and when $i = 0$ it is $\text{Hom}_S(A, \text{Hom}_R(S, Q^j)) \cong \text{Hom}_R(A, Q^j)$. Since the ${}_{\text{hor}}^{1}E$ differential is induced from the differential in Q, we see that

$${}_{\text{hor}}^{2}E^{j,i} = \begin{cases} H^j(\text{Hom}_R(A, Q_*)) = \text{Ext}_R^j(A, B) & \text{for } i = 0 \\ 0 \quad \text{for } i > 0. \end{cases}$$

Thus, as in the last example, the spectral sequence degenerates at the ${}^2E$ term, so $H^j(\text{tot}(\text{Hom}_S(P, \text{Hom}_R(S, Q))) \cong \text{Ext}_R^j(A, B)$.

However, the vertical spectral sequence does not degenerate in this case! We have

$$_{\text{vert}}^{\ 1}E^{i,j} = H^j(\text{Hom}_S(P_i, \text{Hom}_R(S, Q^*)),$$

and since P_i is free over S, this may be written as

$$\text{Hom}_S(P_i, H^j(\text{Hom}_R(S, Q^*))) = \text{Hom}(P_i, \text{Ext}_R^j(S, B)).$$

The $_{\text{vert}}^{\ 1}E$ differential is the map induced by the differential of P, and thus the ${}^2E$ term has the form

$$_{\text{vert}}^{\ 2}E^{i,j} = \text{Ext}_S^i(A, \text{Ext}_R^j(S, B)).$$

The ${}^2E$ differential d_2 maps this term to $\text{Ext}_S^{i+2}(A, \text{Ext}_R^{j-1}(S, B)) = {}_{\text{vert}}^{\ 2}E^{i+2,j-1}$. Since ${}_{\text{hor}}E$ has the same limit as ${}_{\text{vert}}E$, we get

$$_{\text{vert}}^{\ 2}E = \text{Ext}_S(A, \text{Ext}_R(S, B)) \Rightarrow \text{Ext}_R(A, B)$$

as required.

The change of rings spectral sequence is a special case of the **spectral sequence of a composite functor**; we have only used the fact that $\text{Hom}_R(A, B)$ is the composite of the functor $\text{Hom}_S(A, -)$ with the functor $\text{Hom}_R(S, -)$ and the fact that the functor $\text{Hom}_R(S, -)$ takes injectives to injectives. The general construction plays an important role in algebraic geometry, beginning with the Leray spectral sequence. See Exercise A3.50 below.

A3.13.5 Exact Sequence of Terms of Low Degree

In general, the relation between the ${}^rE$ term and the ${}^\infty E$ term of a spectral sequence is somewhat tenuous, but there is often a simple relation between the $H_k(\text{tot}\, F)$ and some of the ${}^2E^{p,q}$. For the sake of definiteness we treat the vertical spectral sequence of a third-quadrant double complex F only. Of course, similar remarks will hold for the horizontal spectral sequence, and also for a first-quadrant double complex; they may be extended to other ${}^rE$ as well (see Exercise A3.48).

Proposition A3.25 (5-term exact sequence). *If $F^{i,j}$ is a third-quadrant double complex, then writing H_i for H^{-i} (tot F), and E for ${}_{\text{vert}}E$, we have*

a. $H_0 \cong {}^2E^{0,0}$.

b. For every i there is a pair of natural maps

$$^2E^{0,-i} \xrightarrow{\iota} H_i \xrightarrow{\kappa} {}^2E^{-i,0}.$$

c. There is a 5-term exact sequence

$$H_2 \xrightarrow{\kappa} {}^2E^{-2,0} \xrightarrow{d_2} {}^2E^{0,-1} \xrightarrow{\iota} H_1 \xrightarrow{\kappa} {}^2E^{-1,0} \to 0,$$

where d_2 is the differential of the spectral sequence.

Proof. We use the fact that E converges to $H(\operatorname{tot} F)$, together with the fact that ${}^2E^{p,q} = 0$ for $p > 0$ and for $q > 0$. For example, to prove part c, look at Figure A3.17 where we have shown some of the ${}^2E$ differentials.

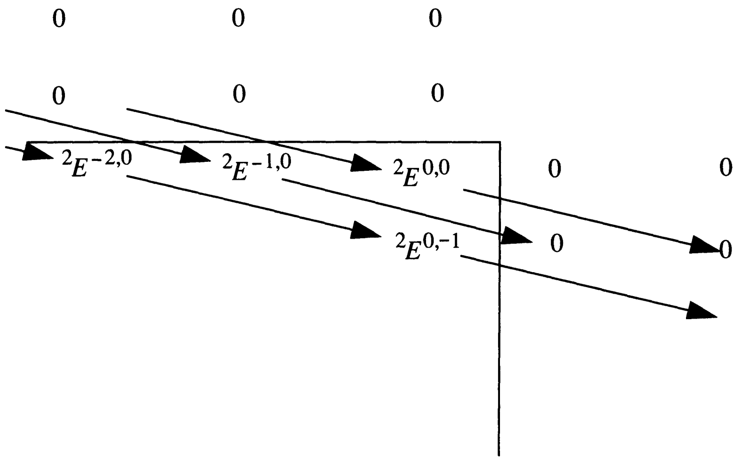


FIGURE A3.17.

Because the terms outside the third quadrant are 0, we see that

$$ {}^2E^{-1,0} = {}^\infty E^{-1,0} = H_1/(H_1)^0, $$

while

$$ \operatorname{coker} d_2 = {}^3E^{0,-1} = {}^\infty E^{0,-1} = (H_1)^0 $$

and

$$ \ker d_2 = {}^3E^{-2,0} = {}^\infty E^{-2,0} = H_2/(H_2)^{-1}. $$

Putting these facts together, we get the five-term sequence. The other parts are similar, but even easier. □

A3.13.6 Exercises on Spectral Sequences

Exercise A3.35: Check the exactness of the derived couple in Proposition A3.20. Check the formulas for ${}^{r+1}Z$ and ${}^{r+1}B$.

Exercise A3.36: Let (F, d) be any differential module, and filter F by

$$ F^0 := F \supset F^1 := \ker d \supset F^2 := \operatorname{im} d \supset F^3 := 0. $$

Writing $({}^rE, d_r)$ for the associated spectral sequence, show that ${}^3E = {}^\infty E = HF$.

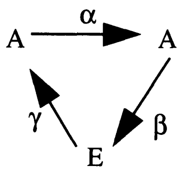
Exercise A3.37: Let $p \in \mathbf{Z}$ be an integer. Explicitly construct the Bockstein spectral sequence associated with the complex

$$F : 0 \to \mathbf{Z}^2 \xrightarrow{(p,0)} \mathbf{Z} \to 0$$

with respect to the endomorphism that is multiplication by p; that is, compute all the ${}^rE$ and d_r, and compute ${}^\infty E$.

Exercise A3.38:

a. Show that for the spectral sequence of the exact couple



there are short exact sequences

$$0 \to A/(\operatorname{im}\alpha + \ker\alpha^r) \to {}^rE \to (\ker\alpha) \cap (\operatorname{im}\alpha^r) \to 0,$$

by showing that the left- and right-hand terms are the images of the appropriate maps in the rth derived exact couple.

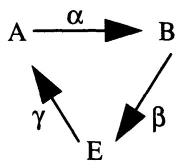
b. Show that if $\ker\alpha^{r+1} = \ker\alpha^r$ for some r, then the spectral sequence collapses at ${}^rE$ and ${}^rE = {}^\infty E = A/(\operatorname{im}\alpha + \ker\alpha^r)$.

c. Show that if A is a finitely generated module and the ground ring is Noetherian, then for some r the condition of part b is satisfied. Give a version that holds for the spectral sequence of a monomorphism $\alpha : F \to F$ of a (not necessarily finite) complex F of finitely generated modules over a Noetherian ring.

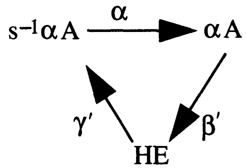
d. If F is a finite complex of finitely generated, torsion-free Abelian groups and p is an integer, then show that the Bockstein spectral sequence for p on F (that is, the spectral sequence for the endomorphism of F that is multiplication by p) has limit $HF/(T + pHF)$, where T is the p-torsion submodule of HF (the set of elements killed by some power of p). For what r is ${}^rE$ equal to ${}^\infty E$?

e. Generalize the argument in the case where F is an infinite complex of finitely generated, torsion-free Abelian groups; again, show that the Bockstein spectral sequence for p has limit $HF/(T + pHF)$.

Exercise A3.39: Generalize the construction of the spectral sequence of an exact couple to the following: Suppose we are given an exact triangle



together with an epimorphism $s : A \to B$. Define a differential $d : E \to E$ by $d = \beta s \gamma$. Show that there is a "derived triangle"



where $\beta'(\alpha a)$ is the class of $\beta s a$, and α', γ' are induced from α, γ; there is also a natural "derived epimorphism" $s : s^{-1}\alpha A \to \alpha A$. Thus the process may be repeated, and we get a spectral sequence.

Exercise A3.40: Let R be the local ring $k[x,y]_{(x,y)}$. Work out all the terms and differentials of the spectral sequence

$$\mathrm{Tor}^{k[x,y]}(k[x,y]/x^2, k[x,y]/xy) \Rightarrow \mathrm{Tor}^R(R/x^2, R/xy + y^3)$$

of Corollary A3.23b.

Exercise A3.41 (Comparison Theorem): Suppose that $F \cdots \supset F^p \supset \cdots$ and $G \cdots \supset G^p \supset \cdots$ are filtered complexes, and that $\alpha : F \to G$ is a morphism of filtered complexes—that is, a morphism of complexes carrying F^p into G^p. Writing ${}^rE(F)$ and ${}^rE(G)$ for the associated spectral sequences, show that there are induced maps ${}^rE(F) \to^r E(G)$ for every r. Show that if one of these maps is an isomorphism, and the spectral sequence converges, then α induces an isomorphism on homology, $H(F) \cong H(G)$.

Exercise A3.42: Let R be a ring, and let $\mathfrak{m}$ be any ideal of R. Recall from Chapter 5 that a filtration

$$\cdots \supset G^p \supset G^{p+1} \supset \cdots$$

of an R-module G is called **$\mathfrak{m}$-stable** if

i. $\mathfrak{m}G^p \subset G^{p+1}$ for all p; and

ii. $\mathfrak{m}G^p = G^{p+1}$ for all sufficiently large p.

In Chapter 5 it is shown that if condition i is satisfied then $\operatorname{gr} G$ is naturally a module over the ring $\operatorname{gr}_{\mathfrak{m}} R$, and that if G is a finitely generated module and both conditions are satisfied, then $\operatorname{gr} G$ is a finitely generated $(\operatorname{gr}_{\mathfrak{m}} R)$-module. Show conversely that if condition i is satisfied and $\operatorname{gr} G$ is a finitely generated $(\operatorname{gr}_{\mathfrak{m}} R)$-module, then ii holds.

Assume that $(R, \mathfrak{m})$ is a local Noetherian ring, and that G is a finitely generated module with $\mathfrak{m}$-stable filtration

$$G \supset \cdots \supset G^p \supset G^{p+1} \supset \cdots .$$

a. If the associated graded module $\operatorname{gr} G = \oplus G^p/G^{p+1}$ is zero, then G is zero.

b. If F is any submodule of G, then the filtration of F by $F^p := F \cap G^p$ is $\mathfrak{m}$-stable. Similarly, the filtration of G/F by $(G/F)^p := (F^p + G)/G$ is $\mathfrak{m}$-stable.

c. Suppose that $(R, \mathfrak{m})$ is a local Noetherian ring. If

$$G: \ \cdots \to G_q \to G_{q-1} \to \cdots$$

is a filtered complex of finitely generated R-modules such that the filtration on each G_q is $\mathfrak{m}$-stable, show that the induced filtration on the homology $H_i G$ is also $\mathfrak{m}$-stable. Prove that the spectral sequence of the filtered complex G converges to HG, that is, $H(\operatorname{gr} G) \Rightarrow HG$.

Exercise A3.43: Prove assertion a of Corollary A3.23 by giving an $\mathfrak{m}$-stable filtration of the Koszul complex $K(x_1, \ldots, x_r)$ as follows: Let δ_i be the degree of the leading form x_i^* of x_i, that is, δ_i is the largest integer δ such that $x_i \in \mathfrak{m}^\delta$. In the Koszul complex, the ith free module may be written as $\wedge^i R^r$. This module has a basic consisting of elements of the form $e_{j_1} \wedge \cdots \wedge e_{j_i}$, where e_j is the basis vector of $\wedge^1 R^r = R^r$ that maps to x_j in R. We filter $\wedge^i R^r$ by submodules

$$(\wedge^i R^r)^p = \oplus (\mathrm{R} e_{j_1} \wedge \cdots \wedge e_{j_i})^p,$$

where $(Re_{j_1} \wedge \cdots \wedge e_{j_i})^p = \mathfrak{m}^{p - \Sigma_t \delta_{j_t}} (Re_{j_1} \wedge \cdots \wedge e_{j_i})$; here $\mathfrak{m}^k$ is interpreted as R for $k \leq 0$.

Show that the filtration $\cdots \supset (\wedge^i R^r)^p \supset (\wedge^i R^r)^{p+1} \supset \cdots$ of $\wedge^i R^r$ is $\mathfrak{m}$-stable, and that with this filtration

$$\operatorname{gr} K(x_1, \ldots, x_r) = K(x_1^*, \ldots, x_r^*),$$

the Koszul complex of the leading forms of the x_i, over the ring $\operatorname{gr}_{\mathfrak{m}} R$. Now deduce assertion a of the corollary from the convergence of the associated spectral sequence

$$H(K(x_1^*, \ldots, x_r^*)) \Rightarrow \operatorname{gr} H(K(x_1, \ldots, x_r)).$$

Exercise A3.44: Prove the assertion of part b of Corollary A3.23 as follows.

a.* First find a free resolution G of M and an $\mathfrak{m}$-stable filtration $\cdots \supset G^p \supset G^{p+1} \supset \cdots$ of it such that the associated graded complex is a free resolution of $\operatorname{gr}_{\mathfrak{m}} M$ over $\operatorname{gr}_{\mathfrak{m}} R$.

b. Define a filtration on the complex $G \otimes N$ by taking

$$(G \otimes N)^p = \text{image in } G \otimes N \text{ of } G^p \otimes N.$$

Show that with respect to this filtration, $\operatorname{gr}(G \otimes N) = \operatorname{gr} G \otimes \operatorname{gr} N$. Now consider the spectral sequence of the filtered complex $G \otimes N$.

Exercise A3.45 (More Change-of-Rings Spectral Sequences): Suppose that $R \to S$ is a homomorphism of rings, and A is an S-module, B an R-module.

a. Show that there is a spectral sequence whose ${}^2E$ term is $\operatorname{Ext}_S(\operatorname{Tor}^R(S, B), A)$, and that converges to $\operatorname{Ext}_R(B, A)$.

b. Similarly, show that there is a spectral sequence

$$\operatorname{Tor}^S_q(\operatorname{Tor}^R_p(S, B), A) \Rightarrow_p \operatorname{Tor}^R_{p+q}(B, A).$$

Exercise A3.46 (The Two-Row and Two-Column Cases):

a. Let F be a double complex whose vertical spectral sequence $E =_{\text{vert}} E$ converges to $H := H(\operatorname{tot} F)$. Suppose that for some r only two columns of ${}^rE$ are nonzero—that is, suppose that the ${}^rE^{p,q}$ are nonzero for only two distinct values of p, say $p = s$ and $p = t$, with $s > t$. Show that there is a long exact sequence

$$\cdots \to {}^rE^{s,i-s} \to H^i \to {}^rE^{t,i-t} \xrightarrow{\delta} {}^rE^{s,i-s+1} \to H^{i+1} \to \cdots,$$

where $\delta = d_{s-t}$ if $r \leq s - t$, and $\delta = 0$ if $r > s - t$.

b. A similar result holds if the ${}^rE^{p,q}$ are nonzero for only two values of q. Apply this to the change-of-rings spectral sequence in the text: For example, assume that $\operatorname{Ext}^j_R(S, B) = 0$ unless $j = s$ or $j = s + 1$, and derive the isomorphism

$$\operatorname{Ext}^s_R(A, B) \cong \operatorname{Hom}_S(A, \operatorname{Ext}^s_R(S, B))$$

and the long exact sequence

$$0 \to \operatorname{Ext}^1_S(A, \operatorname{Ext}^s_R(S, B)) \to \operatorname{Ext}^{s+1}_R(A, B) \to \operatorname{Hom}_S(A, \operatorname{Ext}^{s+1}_R(S, B)) \to$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\operatorname{Ext}^u_S(A, \operatorname{Ext}^s_R(S, B)) \to \operatorname{Ext}^{s+u}_R(A, B) \to \operatorname{Ext}^{u-1}_S(A, \operatorname{Ext}^{s+1}_R(S, B))$$
$$\to \operatorname{Ext}^{u+1}_S(A, \operatorname{Ext}^s_R(S, B)) \to \cdots .$$

c.* Suppose that R is a regular ring of dimension d (for example, a polynomial ring in d variables over a field) and $S = R/I$ is a two-dimensional domain (for example the homogeneous coordinate ring of an irreducible projective curve). Let $B = R$, and let A be any S-module. Show that part b above applies, with $s = d - 2$.

Exercise A3.47: Suppose that $R \to S \twoheadrightarrow k$ are maps of rings. Using the change-of-rings spectral sequence of Exercise A3.45b, show that there is a five-term exact sequence

$$\operatorname{Tor}_2^R(k,k) \to \operatorname{Tor}_2^S(k,k) \to \operatorname{Tor}_1^R(S,k) \to \operatorname{Tor}_1^R(k,k) \to \operatorname{Tor}_1^S(k,k) \to 0.$$

This sequence is particularly interesting when R is a local ring, S is a factor ring of R, and k is the residue field of R and S. Interpret the sequence in this case in terms of minimal free resolutions. See, for example, Gulliksen and Levin [1969] for information on the resolution of the residue class field of a local ring.

Exercise A3.48: Find explicit analogues and generalizations for Proposition A3.25 for all ${}^rE$, for horizontal spectral sequences, and for the case where $F^{i,j} = 0$ for $i < 0$ and $j < 0$.

Exercise A3.49 Resolutions of complexes:* If

$$F : 0 \to F^0 \to F^1 \to F^2 \to \cdots$$

is a complex of modules, show that there is a double complex $I^{j,k}$ and maps

$$\begin{array}{ccccccccc}
 & & \cdots & & \cdots & & \cdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & I^{0,2} & \to & I^{1,2} & \to & I^{2.2} & \to & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & I^{0,1} & \to & I^{1,1} & \to & I^{2,1} & \to & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & I^{0,0} & \to & I^{1,0} & \to & I^{2,0} & \to & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & F^0 & \to & F^1 & \to & F^2 & \to & \cdots
\end{array}$$

such that

i. Each column $I^{j,0} \to I^{j,1} \to I^{j,2} \to \cdots$ is an injective resolution of F_j.

ii. In the rows, the kernel of each $I^{j,k} \to I^{j+1,k}$ is an injective summand of $I^{j,k}$, and thus the image of $I^{j,k} \to I^{j+1,k}$ and the homology of $I^{j-1,k} \to I^{j,k} \to I^{j+1,k}$, which is ${}_{\text{hor}}^{\ 1}E^{j,k}$, are injective modules.

iii. The spectral sequence ${}_{\mathrm{hor}}E$ degenerates at ${}^2E$ to $H(F)$; that is, the term ${}_{\mathrm{hor}}^{\ 1}E$ of the spectral sequence, with differential d_1 induced by the vertical maps in the diagram above, forms injective resolutions

$$0 \to H^j(F) \to {}_{\mathrm{hor}}^{\ 1}E^{j,0} \to {}_{\mathrm{hor}}^{\ 1}E^{j,1} \to {}_{\mathrm{hor}}^{\ 1}E^{j,2} \to \cdots$$

of the homology of F.

Such a double complex is called an injective resolution of the complex F.

Exercise A3.50 (Grothendieck's spectral sequence of a composite functor): Suppose that $\mathcal{A}$ and $\mathcal{B}$ are categories of modules over some rings and that $\mathcal{F} : \mathcal{A} \to \mathcal{B}$ and $\mathcal{G} : \mathcal{B} \to \mathcal{C}$ are left-exact functors. What is the relation between the derived functors $R^i\mathcal{F}, R^i\mathcal{G}$, and $R^i(\mathcal{G}\mathcal{F})$? Under favorable circumstances, it is given by a spectral sequence. Prove this as follows:

a. We say that an object B of $\mathcal{B}$ is $\mathcal{G}$-acyclic if $R^i\mathcal{G}(B) = 0$ for all $i > 0$. Show that if B is any object of $\mathcal{B}$ and $0 \to B \to B^0 \to B^1 \to \cdots$ is an exact sequence of objects of $\mathcal{B}$ with each B^i $\mathcal{G}$-acyclic, then

$$R^i\mathcal{G}(B) = H^i(0 \to \mathcal{G}B^0 \to \mathcal{G}B^1 \to \mathcal{G}B^2 \to \cdots).$$

b.* Now suppose that A has a resolution by $\mathcal{F}$-acyclic objects that are carried by $\mathcal{F}$ to $\mathcal{G}$-acyclic objects. Show that there is a spectral sequence

$${}^2E^{p,q} = R^p\mathcal{F}(R^q\mathcal{G}(A)) \Rightarrow_p R^{p+q}(\mathcal{F}\mathcal{G})(A).$$

c. Show that the change-of-rings spectral sequence given in the text is of this form, where the composite functor is

$$\mathrm{Hom}_S(A, \mathrm{Hom}_R(S, -)) = \mathrm{Hom}_R(A, -).$$

d.* If you know enough about sheaves, derive the Leray spectral sequence $H^i(R^j\pi_*(A)) \Rightarrow H^{i+j}(A)$ for a sheaf of Abelian groups A on a topological space X and a continuous map of spaces $\pi : X \to Y$.

A3.14 Derived Categories

> ... le manque de fondements adéquats d'Algebre Homologique m'avait empêché.... Cette lacune de fondements est sur le point d'être comblée par la thèse de Verdier....
>
> [... the lack of an adequate foundation for homological algebra hindered me.... This gap in the foundations has just been filled by the thesis of Verdier... .]
>
> —Alexandre Grothendieck, 1963

We have given a somewhat primitive view of derived functors simply as things constructed from projective or injective resolutions. Various more axiomatic definitions have been used, but the most complete and powerful seems to be Verdier's formulation by means of his notion of the **derived category** [1977]. We give a very brief sketch of the derived category and the picture of derived functors to which it leads, in the hope that this will help orient the reader. More complete pictures may be found in Hartshorne [1966b, Chapter I] Iversen [1986], Grivel [1987], or Lipman [1995].

As we have seen, the central idea in homological algebra is to replace a module by a projective resolution or an injective resolution: for simplicity we shall stick with projective resolutions for this description, and leave the dualization to the reader. There are two desiderata addressed by the construction of the derived category: First, one would like the association of a module to one of its projective resolutions to be a functor of some kind. Second, one would like to be able to replace a module, or a complex of modules, with a complex of projective modules having the same homology, as in some ways these are easier to manipulate. This leads to a construction in two steps, which we now explain. We shall ignore some set-theoretic points (coming for example from the fact that the "set of all modules" is not a set) that would form a part of a careful treatment.

A3.14.1 Step One: The Homotopy Category of Complexes

The association of a module to its projective resolution is not a functor, because projective resolutions are not unique, and neither are the maps induced on projective resolutions by maps of modules. The first of these problems is easy to cure: We simply choose a fixed projective resolution $P(M)$ for each module M (other, more canonical solutions would be to make a "canonical projective resolution", with each module free on the elements of the kernel of the map before; or to take some direct limit over all projective resolutions). Unfortunately, the nonuniqueness of maps induced on projective resolutions keeps P from being a functor. However, we have already seen that every map of modules lifts to a map of projective resolutions that is unique up to homotopy. Thus P becomes a functor from the category $\mathcal{M}$ of R-modules to the category $K(\mathcal{M})$ whose objects are complexes of R-modules and whose morphisms are the homotopy classes of maps between complexes.

Because we would like to have projective resolutions for any object in our category, and because it is not so clear how to make projective resolutions for unbounded complexes, we restrict ourselves at this point to the category $K^+(\mathcal{M})$ of "bounded-below complexes," that is complexes

$$F:\ \cdots \to F_{i+1} \to F_i \to \cdots$$

with $F_i = 0$ for $i \ll 0$. See Exercise A3.53 for the meaning of projective resolutions in this setting. (Recent developments suggest that there are

good resolutions for unbounded complexes too—see Avramov and Halperin [1986]. In any case, a more thorough treatment of derived categories would contain parallel constructions with bounded-below complexes, unbounded complexes, and bounded-above complexes, the last for the purpose of using injective resolutions and constructing right derived functors. We shall systematically ignore all but the first of these.) Because homotopic maps induce the same map on homology, one still can speak of the "nth homology module" $H_i(X)$ of an object X of $K(\mathcal{M})$, even though one cannot speak of the "term of degree n" in X.

Now the category $K^+(\mathcal{M})$ is no longer an Abelian category. For example, if F and G are the complexes of Abelian groups

$$F: \cdots 0 \to 0 \to \mathbf{Z} \to 0 \to 0 \to \cdots,$$
$$G: \cdots 0 \to 0 \to \mathbf{Z}/(2) \to 0 \to 0 \to \cdots,$$

and $\pi : F \to G$ is the natural map of complexes mapping $\mathbf{Z}$ onto $\mathbf{Z}/(2)$, then no image for π exists in $K^+(\mathcal{M})$; see Exercise A3.51.

Because $K^+(\mathcal{M})$ is not Abelian, we cannot speak of exact sequences in this category. However, the category $K^+(\mathcal{M})$ has a new structure, called a **triangulation**, which can be used as a substitute for exact sequences. First, we have a "translation functor" T on complexes that takes a complex F to the complex $F[-1]$. Given a translation functor T on a category, a triangulation is a distinguished collection of diagrams of the form

$$A \to B \to C \to TA,$$

satisfying certain axioms, which we shall not state. In the case of the category $K^+(\mathcal{M})$, we may take the triangles to be the diagrams made from the mapping cones of maps $\alpha : A \to B$ of complexes, that is the diagrams of the form

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} M(\alpha) \xrightarrow{\gamma} TA,$$

where for simplicity we have written α for the homotopy class of α, and β and γ are the homotopy classes of the standard inclusion of B in $M(\alpha)$ and the projection of $M(\alpha)$ to $A[-1] = TA$, respectively. If we apply the homology functor to such a triangle, we get a long exact sequence in homology, as explained in the section on mapping cones.

The reason for making this choice instead of taking the triangles to be (or at least include) the short exact sequences of complexes is that, with the definition above, any additive functor on the category of modules induces, in an obvious way, a functor on $K^+(\mathcal{M})$ that preserves triangles.

A3.14.2 Step Two: The Derived Category

Following our outline, we wish to be able to replace any complex by a projective complex with the same homology. The construction of projective

resolutions of complexes, Exercise A3.53, shows that given any bounded-below complex F there is a bounded-below complex F' of projective modules and a map $F' \to F$ of complexes that induces an isomorphism on homology. Thus we may attain our goal by formally inverting every morphism that is an isomorphism on homology—such a morphism is called a **quasi-isomorphism**, or **quism**.

Now, quite generally, given a category $\mathcal{A}$ and a set $\mathcal{S}$ of morphisms, there is a universal solution to the problem of finding a category $\mathcal{B}$ and a functor $\mathcal{A} \to \mathcal{B}$ taking all the elements of $\mathcal{S}$ to isomorphisms; the resulting category $\mathcal{B}$ is unique up to equivalence of categories and is called $\mathcal{A}[\mathcal{S}^{-1}]$. The objects of $\mathcal{A}[\mathcal{S}^{-1}]$ may be taken to be the same as the objects of $\mathcal{A}$, and the morphisms are "words" whose letters are morphisms of $\mathcal{A}$ and formal inverses s^{-1} of morphisms $s \in \mathcal{S}$, subject to the condition of composability and the equivalence relation generated by composition in $\mathcal{A}$ and the rule that s^{-1} is inverse to s. (The construction is directly analogous to localization of rings, which is actually the special case where $\mathcal{A}$ is an additive category having just one object X—the ring in question is $\mathrm{Hom}(X, X)$.) However, these localized categories are in general quite awkward. For example, there may be no simple criterion to tell whether two morphisms from $\mathcal{A}$ become equal in $\mathcal{B}$. (The same phenomenon occurs in the special case of localizations of general noncommutative rings. In logical terms, the "word problem" may be recursively insoluble.)

In the case of the category $K^+(\mathcal{M})$ we are lucky (the recognition of this luck seems to have been one of Verdier's fundamental insights): The localization of $K^+(\mathcal{M})$ with respect to the set of quasi-isormorphisms has a nice form. The fundamental point is that the maps in the localized category can all be represented in the form $\alpha^{-1}a$, where a is a morphism of $K^+(\mathcal{M})$ and α is a quasi-isomorphism, so that we have a sort of "calculus of fractions." The crucial point that must be checked is that we can rewrite any composition $b\beta^{-1}$ with β a quism in the form $\alpha^{-1}a$, with α a quism. Rewriting this without using inverses, one must check that given a map $a : B \to C$ and a quasi-isomorphism $\alpha : B \to A$, there exists, for some complex B', a quasi-isomorphism $\beta : C \to B'$ and a map $b : A \to B'$ such that $b\alpha = \beta a$ (see, for example, Hartshorne [1977, p. 30].)

We now define the **derived category** $D^+(\mathcal{M})$ to be the category $K^+(\mathcal{M})$ with the quasi-isomorphisms formally inverted. The objects of $D^+(\mathcal{M})$ are complexes, and the maps are things of the form $a\alpha^{-1} : A \to C$, where $a : B \to C$ is a morphism and $\alpha : B \to A$ is a quasi-isomorphism in $K^+(\mathcal{M})$, modulo an equivalence relation effectively saying that one can cancel a quasi-isomorphism from a product. We write $P : K^+(\mathcal{M}) \to D^+(\mathcal{M})$ for the localization functor.

The derived category inherits from $K^+(\mathcal{M})$ the structure of a triangulated category: Since the translation functor on $K^+(\mathcal{M})$ preserves quasi-isomorphisms, it induces a functor, called again translation, on $D^+(\mathcal{M})$, and we take as a triangle anything quasi-isomorphic to a triangle in $K^+(\mathcal{M})$.

It is interesting to note that any exact sequences of complexes becomes a triangle of $D^+(\mathcal{M})$; this follows from Exercise A3.33.

Just as using the localization of a ring is conceptually simpler than working in the original ring but "as if" elements of a multiplicative system were invertible, working in the derived category has proved simpler for certain applications than working with the category of complexes directly. However, every quasi-isomorphism between bounded-below projective complexes is actually a homotopy equivalence, so that if we define $\mathcal{P}$ to be the category of projective R-modules, the derived category may be described simply as $K^+(\mathcal{P})$; see Exercise A3.54.

With these ideas in place we can describe left-derived functors (for right-derived functors one would use bounded-above complexes and injective resolutions). If F is an additive functor from $\mathcal{M}$ to $\mathcal{M}$, say, then as we have already noted, F induces a functor that we may call $K(F) : K(\mathcal{M}) \to K(\mathcal{M})$. The left-derived functor LF of F is a functor $LF : D^+(\mathcal{M}) \to D^+(\mathcal{M})$, together with a natural transformation $\eta : LF \circ P \to KF$, which gives the "best possible approximation to KF" in the sense that for any functor $G : D^+(\mathcal{M}) \to D^+(\mathcal{M})$ and natural transformation $\nu : G \circ P \to KF$, there is a unique map $G \to LF$ such ν is the composite $G \circ P \to LF \circ P \to F$. The old derived functors L_iF are obtained by composing LF with the "ith homology functor" $H_i : D^+(\mathcal{M}) \to \mathcal{M}$.

The first hint of the simplification that is obtained by all this comes when one considers composite functors. Previously we saw that under good conditions the derived functors of a composite functor fit into a spectral sequence (we did this in the text in the context of "change of rings" and in the exercises in general). But in terms of the derived category, the derived functor of a composite functor is (under the same favorable circumstances) simply the composition of derived functors! For example, if S is an R-algebra, M is an S-module,

$$F = S \otimes_R - : R - \text{modules} \;\to S - \text{modules},$$

and

$$G = M \otimes_S - : S - \text{modules} \;\to S - \text{modules},$$

then the spectral sequence

$$L_iG(L_jF) = \operatorname{Tor}_i^S(M, \operatorname{Tor}_j^R(S,N)) \Rightarrow \operatorname{Tor}_{i+j}^R(M,N) = L_{i+j}(G \circ F)(N),$$

where N is an R-module, is replaced by the much simpler

$$LG \circ LF = L(G \circ F).$$

When there are many functors and compositions around, this simplification can be decisive. Of course, when one wants to make computations one must fall back to the more concrete language of spectral sequences.

A3.14.3 Exercises on the Derived Category

Exercise A3.51 (The category $K^+(\mathcal{M})$ is not Abelian):* In an Abelian category, every morphism $A \to C$ can be factored into an epimorphism followed by a monomorphism $A \twoheadrightarrow B \rightarrowtail C$. Show that the natural map $\mathbf{Z} \to \mathbf{Z}/(p)$ gives rise to a map of complexes

$$A = \{\cdots \to 0 \to \mathbf{Z} \to 0 \to \cdots\} \to \{\cdots \to 0 \to \mathbf{Z}/(p) \to 0 \to \cdots\} = C$$

that cannot be factored in this way in $K^+(\mathcal{M})$.

Exercise A3.52: If A and B are bounded-below complexes of projective modules, and $\alpha : A \to B$ is a quasi-isomorphism, show that α is a homotopy equivalence.

Exercise A3.53: Let F be a bounded-below complex of R-modules. Imitate Exercise A3.49 to show that there is a bounded-below complex of projective R-modules P and a quasi-isomorphism $P \to F$. Such a P is a projective resolution of F.

Exercise A3.54: Let $K^+(\mathcal{P})$ be the category whose objects are bounded-below complexes of projective R-modules, and whose morphisms are homotopy classes of morphisms of complexes. Define a "projective resolution functor" $K^+(\mathcal{M}) \to K^+(\mathcal{P})$. Show that it sends quasi-isomorphisms to isomorphisms (that is, to homotopy equivalences), and thus induces a functor $D^+(\mathcal{M}) \to K^+(\mathcal{P})$. Show that together with the composite functor $K^+(\mathcal{P}) \to K^+(\mathcal{M}) \to D^+(\mathcal{M})$, this defines an equivalence of categories $D^+(\mathcal{M}) \cong K^+(\mathcal{P})$.

Appendix 4
A Sketch of Local Cohomology

As we have often seen, there is a tight analogy between local and graded rings. We have generally started from things that we proved for the local case and adapted them for the graded case. But the analogy flows in the other direction too. A graded domain R gives rise to a subvariety $X = \operatorname{Proj} R$ of projective space, and a module M over R gives rise to a sheaf $\tilde{M}$ on X. One of the most important tools available in this context is the cohomology $H^*(X, \tilde{M})$. If we take the local-global analogy seriously, we should ask whether there is a good local analogue of this cohomology.

The answer is yes, and the corresponding construction is called local cohomology. We will state some of the most basic definitions and results pertaining to it but omit the proofs for the sake of brevity. The reader can find more information in Grothendieck [1967] and Brodmann and Sharp [1996].

First a general definition: If R is a ring, I an ideal of R, and M an R-module, then we define the zeroeth local cohomology module of M with supports in I to be simply the set of all elements of M which are annihilated by some power of I:

$$H_I^0(M) = \cup_n (0 :_M I^n) = \lim_{n\to\infty} \operatorname{Hom}(R/I^n, M),$$

where $(0 :_M I^n)$ denotes the set of elements of M annihilated by I^n. We define the higher local cohomology groups as the right-derived functors of H_I^0—that is, $H_I^i(M)$ is the ith cohomology module of the complex obtained by applying H_I^0 to an injective resolution of M.

Geometrically, if we think of elements of M as global sections of the sheaf on Spec R associated to M, then the elements of $H_I^0(M)$ are just

the sections with support on the closed subscheme Spec $R/I \subset$ Spec R. It is clear that a similar definition could be made for any closed subscheme of any scheme, and indeed the theory is most naturally developed in this context—see, for example, Grothendieck [1967].

It is easy to see that the functor H_I^0 is left-exact, and so for any short exact sequence of modules we get a long exact sequence in local cohomology. Since local cohomology is the derived functor, it is universal among sequences of functors with this property. On the other hand, the functors

$$\lim_{n\to\infty} \operatorname{Ext}_R^i(R/I^n, M)$$

behave in a similar way, taking short exact sequences to long exact sequences. A careful inspection shows that they have the same universal property as the local cohomology, so they are in fact naturally isomorphic:

$$H_I^i(M) \cong \lim_{n\to\infty} \operatorname{Ext}_R^i(R/I^n, M).$$

Besides $\cup_n(0 :_M I^n)$ and $\lim_{n\to\infty} \operatorname{Hom}(R/I^n, M)$, we can express the zeroeth local cohomology in another way in terms of familiar objects. If $I = (x_1, \ldots, x_s)$, then the elements of M annihilated by some power of I are the same as the elements annihilated by some power of each of the x_i. Hence we have

$$H_I^0(M) = \lim_{n\to\infty} H^0(M \otimes K(x_1^n, \ldots, x_s^n)),$$

where the maps

$$H^0(M \otimes K(x_1^n, \ldots, x_s^n)) \to H^0(M \otimes K(x_1^{n+1}, \ldots, x_s^{n+1}))$$

over which the limit is taken are the inclusions

$$\begin{aligned} H^0(M \otimes K(x_1^n, \ldots, x_s^n)) &= (0 :_M (x_1^n, \ldots, x_s^n)) \subset (0 :_M (x_1^{n+1}, \ldots, x_s^{n+1})) \\ &= H^0(M \otimes K(x_1^{n+1}, \ldots, x_s^{n+1})). \end{aligned}$$

Equivalently, and more usefully, we may think of these maps as induced by the natural maps of Koszul complexes

$$K(x_1^n, \ldots, x_s^n) \to K(x_1^{n+1}, \ldots, x_s^{n+1})$$

which in degree 1 are given by the map $f : R^n \to R^n$ multiplying the ith component by x_i, and in degree d are $\wedge^d f$, which acts by multiplying a basis vector $e_{i_1} \wedge \cdots \wedge e_{i_d}$ by $x_{i_1} \cdot \ldots \cdot x_{i_d}$. Thus we may take the limit in each of the Koszul homology groups, and arguing as before we get

$$H_I^i(M) \cong \lim_{n\to\infty} H^i(M \otimes K(x_1^n, \ldots, x_s^n)).$$

A4.1 Local Cohomology and Global Cohomology

The last isomorphism above provides the means to relate local cohomology to the cohomology of coherent sheaves on a projective variety or scheme.

Suppose that R is graded, with maximal ideal P generated by $x_1, \ldots, x_s$, and having degree 0 part, R_0, a field. We write $\tilde{M}$ for the sheaf induced by M on the scheme $\operatorname{Proj} R$.

First a general remark which will help to identify the limit of the Koszul complexes: If we take a sequence of modules $M_n \cong M$, and maps $M_n \to M_{n+1}$ induced by multiplication by some fixed element $a \in R$, then

$$\varinjlim M_n = M[a^{-1}],$$

the localization of M with respect to the multiplicative set generated by a. Of course if a is homogeneous, then the degree 0 part of $M[a^{-1}]$ is the module of sections of $\tilde{M}$ on the open set $a \neq 0$ of $\operatorname{Proj} R$, and $M[a^{-1}]$ itself is the sum over all ν of the global sections of $\tilde{M}(\nu)$ on the open set $a \neq 0$. If $a = x_{i_1} \cdot \ldots \cdot x_{i_d}$, then writing U_i for the open set $x_i \neq 0$, the open set $a \neq 0$ is the intersection $U_{i_1} \cap \cdots \cap U_{i_d}$.

Thus with f as above, the limit of modules isomorphic to $M \otimes \wedge^d R^s$ under the maps induced by $\wedge^d f$ is

$$\lim_{n \to \infty} M \otimes \wedge^d R^s = \oplus_{i_1 \cdots i_d} \sum_\nu H^0(U_{i_1} \cap \cdots \cap U_{i_d}, \tilde{M}(\nu)|_{U_{i_1} \cap \cdots \cap U_{i_d}}).$$

Since taking homology commutes with direct limits over directed sets, we see that the local cohomology of M is the cohomology of the complex

$$0 \to M \to \oplus_i \sum_\nu H^0(U_i, \tilde{M}(\nu)_{|U_i}) \to \ldots$$

$$\to \oplus_{i_1 \cdots i_d} \sum_\nu H^0(U_{i_1} \cap \cdots \cap U_{i_d}, \tilde{M}(\nu)|_{U_{i_1} \cap \cdots \cap U_{i_d}}) \to \ldots,$$

and except for the first term, this is the Čech complex, whose ith homology is the ordinary Čech cohomology H^i $(\operatorname{Proj} R, \tilde{M})$. This shows that local and global cohomologies are related in the following way:

Theorem A4.1. *If M is a graded R-module, then there is a natural exact sequence*

$$0 \to H^0_P(M) \to M \to \sum_\nu H^0(\operatorname{Proj} R, \tilde{M}(\nu)) \to H^1_P(M) \to 0$$

and for every $i > 0$ a natural isomorphism

$$\sum_\nu H^i(\operatorname{Proj} R, \tilde{M}(\nu)) \cong H^{i+1}_P(M),$$

where the sums extend over all positive and negative integers.

One reason that ordinary cohomology is so useful is that each of the $H^i(\operatorname{Proj} R, \tilde{M}(\nu))$ is a finite-dimensional vector space over the field R_0. The local cohomology modules, being infinite-direct sums of these, are not in general finite-dimensional, and in the case where R is local rather than graded, and P is the maximal ideal, they do not break up into such convenient finite pieces. However, if M is finitely generated one can show directly

that the $H^i_P(M)$ are at least Artinian modules. (*Reason:* The ith step in an injective resolution of M consists of a direct sum of injective envelopes $E(R/Q)$ of modules R/Q, with Q a prime ideal, and only finitely many of each $E(R/Q)$ occur. Applying the zeroeth local cohomology functor to one of these $E(R/Q)$ gives 0 unless $Q = P$, in which case it gives $E(R/P)$, so the local cohomology is the homology of a complex of finite direct sums of copies of $E(R/P)$. Since each of these is an Artinian module, the local cohomology is too.)

A4.2 Local Duality

One of the most important results about cohomology is the duality theorem, which for a sheaf $\mathcal{F}$ on a d-dimensional projective space X says

$$H^i(X, \mathcal{F}) \cong \operatorname{Ext}_X^{d-i}(\mathcal{F}, \mathcal{O}_X(-d-1))^*,$$

where * represents Hom into the ground field. Of course if $\mathcal{F}$ is invertible, this degenerates to the more familiar

$$H^i(X, \mathcal{F}) \cong H^{d-i}(X, \mathcal{F}^{-1} \otimes \omega_X)^*.$$

The local form of this is:

Theorem A4.2. *If (S, Q) is a regular local ring of dimension d, and M is a finitely generated R-module, then*

$$H^i_Q(M) \cong \operatorname{Ext}_S^{d-i}(M, R)^*,$$

where * denotes the duality functor $\operatorname{Hom}_R(-, E(R/P))$.

If (R, P) is a factor ring of a regular local ring ring (S, Q), and M is an R-module, then just as in the case of ordinary cohomology it is easy to see that the corresponding local cohomology modules agree:

$$H^i_Q(M) = H^i_P(M),$$

so the theorem reduces all local cohomology questions to questions about Ext modules, at least for rings which are factor rings of regular local rings (virtually every ring of geometrical interest).

A4.3 Depth and Dimension

In particular, one can deduce from Theorem A4.2, Theorem 18.20, and the Auslander-Buchsbaum formula that the functor $\operatorname{Ext}_R^*(-, R)$, and thus also the local cohomology, measures both the depth and the dimension of a module.

Theorem A4.3. *Let (R, P) be a local ring, and let M be a finitely generated R-module. Let $d = \dim M$, and let $\delta = \operatorname{depth}(P, M)$. We have:*

a. $H^i_P(M) = 0$ *for* $i < \delta$ *and for* $i > d$.

b. $H^i_P(M) \neq 0$ *for* $i = \delta$ *and for* $i = d$.

Of course it follows that $\delta \leq d$; in the case where M is a factor ring, this is a consequence of Proposition 18.2, and it can be proved in the same way in general.

Exercise A4.1:

a: Let (R, P) be a local ring such that R_Q is Cohen-Macaulay for every prime ideal $Q \neq P$. Show that $H^i_P(R)$ has finite length for every $i < \dim R$.

b: Let $R = \oplus_{d \geq 0} R_d$ be a Noetherian positively graded ring and suppose that R_0 is a field. Suppose that R_Q is Cohen-Macaulay for each homogeneous prime $Q \neq P$. Show that $R_{Q'}$ is Cohen-Macaulay for every prime Q' of R, homogeneous or not. Then show that $H^i_P(R)$ has finite length for every $i < \dim R$.

Exercise A4.2: Let $R = \oplus_{d \geq 0} R_d$ be a Noetherian positively graded ring and suppose that R_0 is a field. Let $P = \oplus_{d > 0} R_d$ be the maximal homogeneous ideal. Let $R_{(e)} = \oplus_{d \geq 0} R_{de}$ be the eth Veronese subring of R, and let $P_{(e)}$ be its maximal homogeneous ideal.

a: Show that
$$H^i_{P_{(e)}}(R_{(e)}) = H^i_P(R)_{(e)} := \oplus_d H^i_P(R)_{de}.$$

b: Deduce from Exercises A4.1b and A4.2a that if R_Q is Cohen-Macaulay for each homogeneous prime $Q \neq P$, then $H^i_{P_{(e)}}(R_{(e)})$ is concentrated in degree 0 (that is $H^i_{P_{(e)}}(R_{(e)})_d = 0$ for $d \neq 0$) for all sufficiently large integers e and all $i < \dim R$.

Exercise A4.3: A local ring (R, P) is said to be **Buchsbaum** if the natural map $\mathrm{Ext}^i_R(R/P, R) \to \lim_{d \to \infty} \mathrm{Ext}^i_R(R/P^d, R) = H^i_P(R)$ is an isomorphism for every $i < \dim R$. It turns out that this somewhat unappetizing definition leads to a rich and surprising theory; see Stückrad and Vogel [1986]. Show that a sufficiently high Veronese embedding of any projective scheme has Buchsbaum homogeneous coordinate ring as follows. Let $R = \oplus_{d \geq 0} R_d$ be a Noetherian positively graded ring and suppose that R_0 is a field. Suppose that $H^i_P(R)$ is concentrated in degree 0 for $i < d$ as in Exercise A4.2b. Show that R is Buchsbaum.

Appendix 5
Category Theory

A5.1 Categories, Functors, and Natural Transformations

A **category** $\mathcal{C}$ is a collections of **objects** and for each pair of objects A, B a set $\mathrm{Hom}_{\mathcal{C}}(A, B)$ of **morphisms** with a composition law

$$\mathrm{Hom}_{\mathcal{C}}(A,B) \times \mathrm{Hom}_{\mathcal{C}}(B,C) \to \mathrm{Hom}_{\mathcal{C}}(A,C) \qquad (f,g) \mapsto gf$$

and a distinguished element $1_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$ for each object A such that:

a. Composition is associative in the sense that $h(gf) = (hg)f$ whenever both sides are defined.

b. $f1_A = f$ and $1_A g = g$ whenever the compositions are defined.

We usually write $f : A \to B$ to mean $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$, and we say that certain diagrams commute to indicate that certain compositions are equal.

In this book the category that appears (implicitly) most frequently is the category R-Mod, where R is a ring, whose objects are the modules over R and whose morphisms are the homomorphisms of R-modules. Variants include the subcategories of finitely generated or finitely presented modules. Of course the categories Ring (objects are commutative rings, morphisms are ring homomorphisms) and its subcategories of algebras over a fixed ring, affine rings, and so forth, are also important to us.

Note that in the category R-mod, as in the category of sets, the collection of objects is not itself a set. It is not hard to see that one can always restrict

it to be a set in practical applications, and it is sometimes essential to do so. All the methods I know for doing this seem somewhat artificial. One standard device is Grothendieck's idea of "universes"; see Grothendieck [1972, Chapter I].

Categories were defined by Eilenberg and Maclane to unify ideas from group theory and topology, and the language of category theory (though not many specific results) is now used very widely. Maclane [1971] gives an overview. For homological algebra, the setting of **Abelian categories**, defined by Buchsbaum in his thesis and applied by Grothendieck, has become the standard; it has the advantage of including both the categories of modules (such as R-Mod) in commutative algebra and the categories of sheaves in algebraic geometry, as well as many other useful examples. See Freyd's seductive little book [1964] for an introduction.

In this book we have often used the notions of functor and natural transformation and we have occasionally mentioned the notion of adjoint functor. Here we provide a brief introduction. Another useful categorical construction, that of limit and colimit, is taken up, with some important examples, in Appendix 6.

A map between categories is called a **functor:** More explicitly, if $\mathcal{C}$ and $\mathcal{D}$ are categories, then a functor $F : \mathcal{C} \to \mathcal{D}$ is

a. For every object A of $\mathcal{C}$ an object FA of $\mathcal{D}$.

b. For every morphism $f : A \to B$ a morphism $Ff : FA \to FB$ preserving composition and identity elements.

There are morphisms of functors too (so the category of all categories is a "double category" in a certain sense): If $F, G : \mathcal{C} \to \mathcal{D}$ are functors, then a morphism $\alpha : F \to G$, called a **natural transformation,** is, for each object A of $\mathcal{C}$, a morphism $\alpha_A : FA \to GA$ such that whenever $f : A \to B$ is a morphism in $\mathcal{C}$, the diagram

$$\begin{array}{ccc} FA & \xrightarrow{\alpha_A} & GA \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Gf} \\ FB & \xrightarrow[\alpha_B]{} & GB \end{array}$$

commutes. We say that F and G are **isomorphic**, written $F \cong G$, if there are natural transformations $\alpha : F \to G$ and $\beta : G \to F$ whose compositions $\alpha\beta$ and $\beta\alpha$ are the identity natural transformation (that is, $\beta_A\alpha_A = 1_{FA}$ and $\alpha_A\beta_A = 1_{GA}$ for all objects A of $\mathcal{C}$). The following section, on adjoint functors, contains a number of examples.

The functors with which we have most to do, such as the functors F, G : R-Mod $\to$ R-Mod defined by $F(N) = M \otimes_R N$ and $G(N) = \mathrm{Hom}_R(M, N)$ for some fixed module M, as well as the functors made from Tor or Ext, are **additive** in the sense that $F : \mathrm{Hom}(M, N) \to \mathrm{Hom}(F(M), F(N))$ is a homomorphism of Abelian groups.

Additive functors preserve finite direct sums: To say that $M \cong N_1 \oplus N_2$ means that there are inclusion and projection maps $\iota_i : N_i \to M$ and $\pi_i : M \to N_i$ such that $\pi_i \iota_i = 1_{N_i}$ and $\iota_1\pi_1 + \iota_2\pi_2 = 1_M$. These formulas are preserved by any additive functor.

Actually the functors on R-mod that we commonly use have an even stronger property: They are ***R*-linear** in the sense that

$$F : \mathrm{Hom}(M, N) \to \mathrm{Hom}(F(M), F(N)) \text{ is a homomorphism of } R\text{-modules.}$$

For such a functor, taking $M = N$, we see that if $f_M : M \to N$ is multiplication by an element $r \in R$, and F is R-linear, then $F(f_M) = f_{F(M)}$. In particular, if f_M is 0 (that is, $r \in \mathrm{ann}(M)$), then $f_{F(M)} = F(f_M) = F(0) = 0$, so $\mathrm{ann}(F(M)) \supset \mathrm{ann}(M)$.

Isomorphism of categories may be defined as usual, but it is the "wrong" notion in the sense that it does not include many interesting cases. To take a very simple example, consider the category $\mathcal{C}$ of finite-dimensional vector spaces over a fixed field k, and let $\mathcal{D}$ be the "opposite" category: That is, the objects of $\mathcal{D}$ are also finite-dimensional vector spaces, but we define $\mathrm{Hom}_{\mathcal{D}}(V, W) := \mathrm{Hom}_{\mathcal{C}}(W, V)$. Let F be the functor that takes each vector space V in $\mathcal{C}$ to its dual V^*, regarded as an object of $\mathcal{D}$, and each homomorphism $f : V \to W$ to the dual homomorphism $f^* : W^* \to V^*$ regarded as a morphism $f^* : F(V) \to F(W)$. It is easy to see that F is a functor. The same kind of dualization defines a functor $G : \mathcal{D} \to \mathcal{C}$, and the composites GF and FG are each the functor that replaces a vector space by its double dual. Thus GF and FG are "like" the identity functors, and we should clearly regard F and G as showing that $\mathcal{C}$ and $\mathcal{D}$ are "equivalent" categories. But GF and FG are *not* the identity functors! The double dual of a vector space V is a vector space whose elements are linear functionals on the dual of V—this is not the same as elements of V. Thus we have not shown that F and G are isomorphisms, and indeed they are not, since, for example, not every vector space has elements that are linear functionals. We thus make a formal definition that includes the case above:

A functor $F : \mathcal{C} \to \mathcal{D}$ is an **equivalence** of categories if there exists a functor $G : \mathcal{D} \to \mathcal{C}$ such that FG and GF are isomorphic to the identity functors.

A5.2 Adjoint Functors

One of the most useful notions from category theory is that of an adjoint of a functor. The notion generalizes the notion of equivalence just exhibited. We present some of the theory as a sequence of problems.

Definition. *If $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ are functors, then we say that F is a* ***left adjoint*** *for G (equivalently: G is a* ***right adjoint*** *for F) if there is a natural isomorphism $\alpha : \mathrm{Hom}_{\mathcal{A}}(-, G(-)) \cong \mathrm{Hom}_{\mathcal{B}}(F(-), -)$.*

This means that for every pair of objects A of $\mathcal{A}$ and B of $\mathcal{B}$ there is an isomorphism $\alpha_{A,B} : \mathrm{Hom}_{\mathcal{A}}(A, G(B)) \cong \mathrm{Hom}_{\mathcal{B}}(F(A), B)$ such that for every morphism of objects $A \to A'$ in $\mathcal{A}$ and $B \to B'$ in $\mathcal{B}$, the diagram

$$\begin{array}{ccc}
\mathrm{Hom}_{\mathcal{A}}(A', G(B)) & \overset{\alpha_{A',B}}{\cong} & \mathrm{Hom}_{\mathcal{B}}(F(A'), B) \\
{\scriptstyle \mathrm{Hom}_{\mathcal{A}}(\varphi, G(\varphi))}\downarrow & & \downarrow{\scriptstyle \mathrm{Hom}_{\mathcal{A}}(F(\varphi), \psi)} \\
\mathrm{Hom}_{\mathcal{A}}(A', G(B')) & \overset{\alpha_{A,B'}}{\cong} & \mathrm{Hom}_{\mathcal{B}}(F(A), B')
\end{array}$$

commutes, where the vertical maps are induced by φ and ψ. We shall sometimes say that (F, G) is an ***adjoint pair*** *of functors.*

A5.2.1 Uniqueness

Show that any two left adjoints of a functor G are naturally isomorphic; dually, two right adjoints of a functor F are naturally isomorphic. (*Hint:* If F and F' are left adjoints to G, let $\varphi_A : F(A) \to F'(A)$ be the image of the identity map 1_{FA} under the adjointness isomorphisms

$$\mathrm{Hom}_{\mathcal{B}}(FA, FA) \cong \mathrm{Hom}_{\mathcal{A}}(A, GFA) \cong \mathrm{Hom}_{\mathcal{B}}(F'A, FA);$$

and define $\psi_A : FA \to F'A$ similarly. Show that φ and ψ are natural transformations, and show that they are isomorphisms by showing that they are inverse to one another.)

A5.2.2 Some Examples

Pairs of adjoint functors occur very frequently in mathematics. Here are a few examples connected with the subjects of this book:

a. Let A be a ring, and let (A-Alg) be the category of commutative A-algebras. Let (Sets) be the category of sets. Let G : (A-Alg) $\to$ (Sets) be the "forgetful functor" which associates to each A-algebra its underlying set, and let F:(Sets) $\to$ (A-Alg) be the "free algebra" functor taking a set X to the polynomial ring $A[X]$ whose indeterminates are the elements of X. Show that F is a left adjoint of G.

b. Let (A-Mod) be the category of A-modules. Let G : (A-Alg) $\to$ (A-Mod) be the "forgetful functor" which associates to an algebra its underlying module, and let F : (A-Mod) $\to$ (A-Alg) be the symmetric algebra functor, which associates to a module M the algebra $\mathrm{Symm}(M)$. Show again that F is a left adjoint of G. In general, "free" constructions tend to be left adjoints to "forgetful" constructions.

c. Let F be the functor which associates to an A-module M the "graded-commutative" algebra $\wedge M$, the exterior algebra. Find a right adjoint for F.

d. Let G: (Groups) $\to$ (Sets) be the forgetful functor from the category of (not necessarily Abelian) groups to sets. Find a left adjoint of G.

e. Let B be an A-algebra, and let U : (B-Mod) $\to$ (A-Mod) be the "forgetful functor" which takes a B-module to the underlying A-module. Show that the functor which takes an A-module M to the B-module $B \otimes_A M$ is a left adjoint of G. Show that the functor which takes an A-module M to the B-module $\mathrm{Hom}_A(B, M)$ is a right adjoint of U.

f. Let A be a ring and let M be any A-module. The functor $N \mapsto M \otimes_A N$ from (A-Mod) to itself is the left adjoint of the functor $N \mapsto \mathrm{Hom}_A(M, N)$.

g. Let $\mathcal{B}$ be a small category (a category whose objects form a set), and $\mathcal{A}$ any category. Let $U : \mathcal{A} \to \mathcal{F}\mathrm{un}\,(\mathcal{B}, \mathcal{A})$ be the functor which associates to an object A of $\mathcal{A}$ the "constant functor" $C : \mathcal{B} \to \mathcal{A}$, which takes every object of $\mathcal{B}$ to the object A and every morphism of $\mathcal{B}$ to the identity morphism 1_A. Show that $\mathcal{A}$ has colimits over functors from $\mathcal{B}$ iff the functor C has a left adjoint, and that in this case $\varinjlim$ is the left adjoint. Similarly, $\varprojlim$, if it exists, is a right adjoint. (Definitions of $\varinjlim$ and $\varprojlim$ are given in Appendix 6.)

h. Let A be a ring and let $\mathcal{B}$ be the category whose objects are triples consisting of an A-algebra B, a B-module M, and an A-linear derivation $d : B \to M$ and whose morphisms $(B, M, d) \to (B', M', d')$ are pairs (φ, ψ) consisting of a ring homomorphism $\varphi : B \to B'$ and a homomorphism of B-modules $\psi : M \to M'$ (where we regard M' as a B-module by means of ψ) such that $d'\varphi = \psi d$. Let $G : \mathcal{B} \to$ (A-Alg) be the forgetful functor. Show that the functor (A-Alg) $\to \mathcal{B}$ taking B to the triple $(B, \Omega_{B/A}, d_B)$, where d_B is the universal derivation, is a left adjoint to G.

A5.2.3 Another Characterization of Adjoints

If $F : \mathcal{A} \to \mathcal{B}$ is a left adjoint of $G : \mathcal{B} \to \mathcal{A}$, then for each object B of $\mathcal{B}$ the identity morphism $GB \to GB$ gives rise by the isomorphism in the definition of adjointness to a morphism $\varepsilon_B : FGB \to B$, called the counit, and similarly for each object A of $\mathcal{A}$ we get a morphism $\eta_A : A \to GFA$ called the unit of the adjoint pair.

a. Show that ε and η are natural transformations, and that the adjointness isomorphism itself can be reconstructed from them: If $\varphi : A \to$

GB is a map, then the corresponding map $FA \to B$ is obtained by composing $F\varphi : FA \to FGB$ with the counit $\varepsilon_B : FGB \to B$. (Similarly, if $\psi : FA \to B$ is a morphism, then the corresponding morphism $A \to GB$ is the composite of the unit $\eta_A : A \to GFA$ with $G\psi : GFA \to GB$.)

b. Conversely, show that if $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ are functors, and if $\varepsilon : FG \to 1$ and $\eta : 1 \to GF$ are natural transformations with the property that $G\varepsilon \circ \eta G = 1_G$ (that is, for each object B of B the identity transformation $GB \to GB$ is the composite of $\eta_{GB} : GB \to GFGB$ and $G(\varepsilon_B) : GFGB \to GB$) and $\eta F \circ F\varepsilon = 1_F$, then F is left adjoint to G with unit η and counit ε.

A5.2.4 Adjoints and Limits

Sometimes one gets a bonus for noticing that a functor has an adjoint, because this fact forces certain properties on the functor. Here are some useful examples:

a. Show that left-adjoint functors preserve colimits (and dually right-adjoint functors preserve limits) in the sense that if $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ are functors, with F left adjoint to G, and if $D : \mathcal{D} \to \mathcal{B}$ is a diagram, then

$$F(\varinjlim D) \cong \varinjlim FD \text{ (natural isomorphism).}$$

b. Use this and the earlier example 2f to show that if M is an A-module then the functor $M \otimes_A$ - preserves right-exact sequences and direct sums.

c. Use this and Example 2h to give another proof of Theorem 16.8 by identifying colimits in the category $\mathcal{B}$.

d. Use the same idea, with Example 2c, to show that if M = coker $\varphi : P \to Q$ is a presentation of a module M, then

$$\wedge^i M = \operatorname{coker} P \otimes \wedge^{i-1} Q \to \wedge^i Q$$

and that

$$\wedge^i (M \oplus N) \cong \bigoplus_{j+k=i} \wedge^j M \otimes \wedge^k N,$$

useful formulas for studying the exterior powers of modules. Find and prove the similar formulas for symmetric powers using Example 2b.

A5.3 Representable Functors and Yoneda's Lemma

There are functors that can be defined directly from the structure of a category. These are called representable functors. Let $\mathcal{A}$ be a category, and let A be an object of $\mathcal{A}$. The **representable functor** (represented by A) is defined to be the functor $h_A(B) = \mathrm{Hom}_{\mathcal{A}}(A, B)$.

As an example, let k be a commutative ring and let R be a commutative k-algebra. Let F be the functor from R-mod to the category of sets that takes any R-module M into the set $\mathrm{Der}_k(R, M)$ of R-linear derivations. The universal property of $\Omega_{R/k}$ may be expressed, as was done in Chapter 16, by saying that this functor is represented by the module $\Omega_{R/k}$ of Kähler differentials: That is, there is a natural isomorphism

$$\mathrm{Der}_k(R, -) \cong \mathrm{Hom}_R(\Omega_{R/k}, -).$$

Universal properties can all be expressed in terms of representable functors, as well as in terms of adjointness as is done above.

Representable functors were given a large role in algebraic geometry by Grothendieck [1957]. Since his work, schemes are commonly described by describing the functors they represent. Often the functors exist (or are known to exist) much more generally than the schemes, and the problem of proving the existence of the schemes becomes the problem of showing that the functors are representable. We followed a similar path in the beginning of Chapter 16. See Eisenbud and Harris [1992] for some elementary examples from the theory of schemes, and nearly any work by Grothendieck for further examples.

One reason that this procedure is useful is that the representable functors, as functors, reflect the properties of the objects that represent them extremely well. The basic observation in this direction is Yoneda's Lemma. To state it, let $\mathcal{F}$ be the category whose objects are functors from $\mathcal{A}$ to the category of sets and whose morphisms are natural transformations. Let $h : A \to \mathcal{F}$ be the functor that takes each object A of $\mathcal{A}$ to the functor h_A. Yoneda's Lemma shows that the functor h is actually an equivalence between $\mathcal{A}$ and the opposite of the category of representable functors on $\mathcal{A}$.

Lemma A5.1 (Yoneda's Lemma). *Let A and B be objects of the category $\mathcal{A}$. With notation as above,*

$$\mathrm{Hom}_{\mathcal{F}}(h_A, h_B) \cong \mathrm{Hom}_{\mathcal{A}}(B, A)$$

by the map sending a natural transformation $\varphi : h_A \to h_B$ to $\varphi_A(1_A) \in$ $\mathrm{Hom}_{\mathcal{A}}(B, A)$.

Proof. The inverse map sends $f \in \mathrm{Hom}_{\mathcal{A}}(B, A)$ to the natural transformation $\varphi : h_A \to h_B$ taking $g \in \mathrm{Hom}_{\mathcal{A}}(A, C) = h_A(C)$ to $\varphi_A(g) := gf \in \mathrm{Hom}_{\mathcal{A}}(B, C) = h_B(C)$. □

Appendix 6
Limits and Colimits

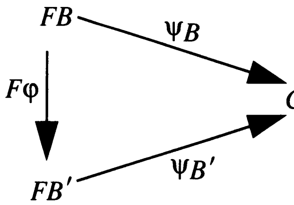
In this section we describe two categorical notions that are useful in commutative algebra, limits and colimits. The two notions are dual to each other, although their realizations are not. Thus for the abstract description, we need only describe one.

The applications we will make involve the category of modules over a ring and the category of algebras over a ring. In the category of algebras the colimit is the deeper notion, as it involves tensor products instead of ordinary products. Also, we will give in this section Govorov and Lazard's famous characterization of flat modules as filtered colimits of free modules. For these reasons we will describe the abstract notion of colimits instead of that of limits.

To get the corresponding definitions and results for abstract limits, the reader need only remove all the "co" prefixes, reverse all the arrows, and interchange the words "source" and "target" in the following discussion. A very important group of applications of the notion of limits is found in Chapter 7, Completions.

Let $\mathcal{A}$ be a category. We define a **diagram** in $\mathcal{A}$, based on $\mathcal{B}$, to be a functor $F : \mathcal{B} \to \mathcal{A}$. Often we will think of $\mathcal{B}$ as a subcategory of $\mathcal{A}$, and we suppress the functor and speak of $\mathcal{B}$ itself as a diagram.

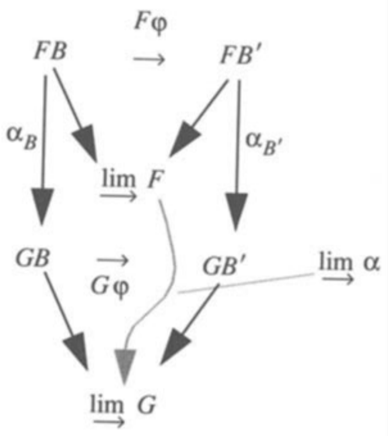
If C is an object of $\mathcal{A}$ and $F : \mathcal{B} \to \mathcal{A}$ is a diagram, we define a **morphism** $\psi : F \to C$ to be a collection $\{\psi_B\}$ consisting of a morphism $\psi_B : FB \to C$ for each object B of $\mathcal{B}$, such that if $\varphi : B \to B'$ is any morphism in $\mathcal{B}$, then the following diagram commutes.



The **colimit** $\varinjlim F$ of F is defined to be an object A of $\mathcal{A}$, and a morphism $\psi : F \to A$, which is **universal** in the sense that for any morphism ψ' from F to an object A', there exists a unique morphism $\gamma : A \to A'$ making all the "obvious" identities $\gamma\psi_B = \psi'_B$ true. Such a colimit, if it exists, is unique up to isomorphism. (*Proof*: If A and A' were colimits, then the universal property would furnish us maps $\gamma : A \to A'$ and $\gamma' : A' \to A$ making the "obvious" identities true. The composite $\gamma'\gamma : A \to A$ would then also be a map making the "obvious" identities true. As the identity map has this same property, the uniqueness statement in the universal property implies that $\gamma'\gamma = 1$. Similarly $\gamma\gamma' = 1$, so γ and γ' are the required isomorphisms.)

In practice, it is often a little artificial to take colimits over a category: The identity morphisms and the compositions of morphisms in $\mathcal{B}$ really play no role, and we may use the same words as above to define the colimit of a set $\mathcal{B}'$ of objects of $\mathcal{A}$ and a set of morphisms between them —this will be the same as the colimit of the subcategory $\mathcal{B}$ generated by $\mathcal{B}'$ (see Exercise A6.9).

The construction $\varinjlim$ itself is a functor in the following way. Let $\mathfrak{Fun}(\mathcal{B}, \mathcal{A})$ be the category of functors from $\mathcal{B}$ to $\mathcal{A}$, whose morphisms are the **natural transformations** of functors. If colimits exist in $\mathcal{A}$, then $\varinjlim$ is a functor from $\mathfrak{Fun}(\mathcal{B}, \mathcal{A})$ to $\mathcal{A}$: Given a natural transformation $\alpha : F \to G$, the diagrams



commute and we define $\varinjlim \alpha : \varinjlim F \to \varinjlim G$ to be the unique map α induced by the vertical maps $FB \to \varinjlim G$.

The following result shows that all colimits can be constructed from two special types of colimits: coproducts and coequalizers. The construction itself, given in the proof, is at least as useful as the existence result.

We define the **coproduct** of a collection of objects $\{B_i\}$ of $\mathcal{A}$, written $\amalg_i B_i$, to be the colimit of the diagram consisting of the objects B_i with no morphisms other than the identity morphisms. We define the coequalizer **coequalizer**(ψ, ψ') of a pair of morphisms

$$C_1 \underset{\psi'}{\overset{\psi}{\rightrightarrows}} C_2$$

to be the colimit of the diagram with objects C_1, C_2, their identity morphisms, and the morphisms ψ, ψ'.

Since we have not required the collection of objects in a category to form a set, colimits of functors from arbitrary categories may fail to exist for somewhat trivial reasons: Most categories (for example, the category of modules over a ring) *don't* contain coproducts of arbitrary collections of objects. A great variety of categories (again including the category of modules over a ring) *do*, however, contain the coproduct of a *set* of objects. Thus it is sensible in discussing colimits to stick to the case of functors from **small** categories, where by a small category we mean one whose collection of objects is a set.

We have:

Theorem A6.1. *If coproducts of sets of objects and coequalizers of pairs of morphisms exist in the category $\mathcal{A}$, then all colimits of functors from small categories exist in $\mathcal{A}$. Further, any functor on $\mathcal{A}$ that preserves coproducts and coequalizers preserves all colimits over small categories.*

Proof. Let $\mathcal{B}$ be a diagram of $\mathcal{A}$. For simplicity, we identify $\mathcal{B}$ with its image in $\mathcal{A}$, and think of it as a subcategory. Let $C_2 := \amalg_{B\in \text{ obj } \mathcal{B}} B$ be the coproduct of all the objects in $\mathcal{B}$. For each morphism φ of $\mathcal{B}$, write source(φ) and target(φ) for the source and target of the map φ. Let $C_1 := \amalg_{\varphi\in \text{ morph } \mathcal{B}}$ source(φ). Thus C_1 is the direct sum of the same objects as C_2, but each object appears in C_1 as many times as there are morphisms originating from it.

We will now define two maps, ψ and ψ', from C_1 to C_2 in such a way that

$$\varinjlim \mathcal{B} = \text{coequalizer}(\psi, \psi').$$

To define a map from C_1 to C_2, it is enough by the universal property of the coproduct to define a map from each of the objects of source(φ) to C_2. On the other hand, from the definition of C_2, there is for each object B of $\mathcal{B}$ a natural "inclusion" map $\iota : B \to C_2$. Thus we may define ψ to be the map induced by the maps

$$\text{source}(\varphi) \xrightarrow{\iota} C_2$$

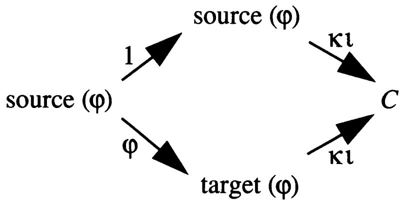
and define ψ' to be the map induced by the composite maps

$$\text{source}(\varphi) \xrightarrow{\varphi} \text{target}(\varphi) \xrightarrow{\iota} C_2.$$

Let $C :=$ coequalizer (ψ, ψ'), and let $\kappa : C_2 \to C$ be the corresponding morphism. For each $B \in \text{obj}\,\mathcal{B}$ we get a composite map

$$B \xrightarrow{\iota} C_2 \xrightarrow{\kappa} C$$

and together these form a morphism $\mathcal{B} \to C$: For each $\varphi \in \text{morph}\,\mathcal{B}$, the diagram



commutes exactly because κ is the coequalizer. Conversely, given any morphism $\mathcal{B} \to C'$, the induced morphism $\kappa' : C_2 \to C'$ will satisfy $\kappa'\psi = \kappa'\psi'$, and thus κ' will factor uniquely through κ, as required.

For the last statement, note that any functor preserving coproducts and coequalizers preserves the construction that we have given for colimits. □

A6.1 Colimits in the Category of Modules

In this subsection we fix a commutative ring A and consider colimits in the category (A-Mod) of modules over A. Limits are treated in Exercises A6.1 and A6.11 (and used extensively in Chapter 7).

Proposition A6.2. *Colimits exist in the category of A-modules. The coproduct is the direct sum, while the coequalizer of a pair of maps is the cokernel of their difference.*

Proof. By the general construction of colimits above, it is enough to prove the second statement. Both parts of this statement follow easily from the definitions. □

Our next goal is Lazard's theorem characterizing flat modules in terms of colimits. The crucial ingredients are the characterization of flat modules by equations from Corollary 6.5 and the notion of a filtered colimit.

To set the stage, we note that if a category $\mathcal{A}$ has colimits, then for any small category $\mathcal{B}$ the functor $\varinjlim : \mathcal{F}\text{un}(\mathcal{B}, \mathcal{A}) \to \mathcal{A}$ is right-exact. This rather abstract statement, which follows directly from the universal property of the colimit (see the section on adjoints and limits, in Appendix 5, Category Theory), has a simple interpretation when $\mathcal{A} = (A\text{–Mod})$: It

means that if F', F, and F'' are functors from $\mathcal{B}$ to $\mathcal{A}$, and if $\alpha : F' \to F$ and $\beta : F \to F''$ are natural transformations such that, for every object B of $\mathcal{B}$,

$$0 \to F'B \xrightarrow{\alpha_B} FB \xrightarrow{\beta_B} F''B \to 0$$

is an exact sequence, then

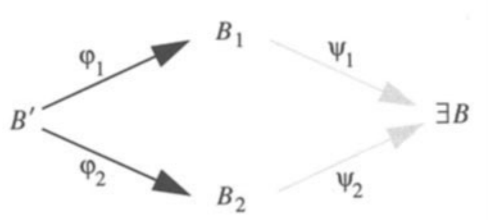
$$\varinjlim F' \xrightarrow{\varinjlim \alpha} \varinjlim F \xrightarrow{\varinjlim \beta} \varinjlim F'' \to 0$$

is a right-exact sequence. What we want is a condition on $\mathcal{B}$ under which it is also left-exact.

Definition (Filtered colimits). *A category $\mathcal{B}$ is said to be **filtered** if:*

a. For any two objects B_1 and B_2 of B there exists an object B of $\mathcal{B}$ with morphisms $B_1 \to B$ and $B_2 \to B$; and

b. For every two morphisms with the same source $\varphi_1 : B' \to B_1, \varphi_2 : B' \to B_2$ there exists an object B and morphisms $\psi_1 : B_1 \to B$ and $\psi_2 : B_2 \to B$ such that $\psi_1\varphi_1 = \psi_2\varphi_2$, as in the following diagram.



A diagram $F : \mathcal{B} \to \mathcal{A}$ is called filtered if $\mathcal{B}$ is filtered, and then we say that $\varinjlim F$ is filtered.

Filtered colimits are like unions of submodules and admit an alternate description.

Proposition A6.3. *If $\mathcal{B}$ is a filtered small subcategory of $(A-\text{Mod})$, then $\varinjlim \mathcal{B}$ is the disjoint union $\cup_{B \in \text{ Obj } \mathcal{B}} B$ (as sets) modulo the equivalence relation $\sim$ defined by:*

$$b_1 \sim b_2 \quad \text{for } b_1 \in B_1 \text{ and } b_2 \in B_2 \text{ iff there exist morphisms}$$
$$\varphi_i : B_i \to B \text{ in } \mathcal{B} \text{ such that } \varphi_1(b_1) = \varphi_2(b_2).$$

Proof. Let $X = \cup_{B \in \text{ Obj } \mathcal{B}} B/\sim$. Because $\mathcal{B}$ is filtered, X has a natural structure of A-module (for example, to add two elements, just map them into a common object of $\mathcal{B}$). The fact that X is the colimit may be seen by checking the universal property directly: Given an A-module C and maps $\psi_i : B_i \to C$, there is an obvious induced map $\psi : \cup_{B \in \text{ Obj } \mathcal{B}} B \to C$, and

if the ψ_i commute with all the maps of $\mathcal{B}$, then ψ respects the equivalence relation $\sim$, and thus descends to a uniquely defined map on X. □

Proposition A6.4. *Filtered colimits preserve exact sequences in the following sense: Let $\mathcal{B}$ be a filtered category and let F', F, and F'' be functors from $\mathcal{B}$ to $\mathcal{A}$. If $\alpha : F' \to F$ and $\beta : F \to F''$ are natural transformations such that, for every object B of $\mathcal{B}$,*

$$0 \to F'B \xrightarrow{\alpha_B} FB \xrightarrow{\beta_B} F''B \to 0$$

is an exact sequence, then

$$0 \to \varinjlim F' \xrightarrow{\varinjlim \alpha} \varinjlim F \xrightarrow{\varinjlim \beta} \varinjlim F'' \to 0$$

is an exact sequence.

Proof. The right-exactness being clear, it is enough to show that $\varinjlim \alpha$ is injective. But this follows at once from Proposition A6.3: If an element of $\cup_{B\in \text{ Obj } \mathcal{B}} F'B/\sim$ goes to zero in $\cup_{B\in \text{Obj}\,\mathcal{B}} FB/\sim$, it is represented by an element b in some $F'B$ such that $\alpha_B(b)$ goes to zero under a map $F\varphi$, for some morphism $\varphi : B \to B'$ of $\mathcal{B}$. But since the diagram

$$\begin{array}{ccc} F'B & \xrightarrow{\alpha_B} & FB \\ {\scriptstyle F'\varphi}\downarrow & & \downarrow{\scriptstyle F\varphi} \\ F'B' & \xrightarrow{\alpha_{B'}} & FB' \end{array}$$

commutes and $\alpha_{B'}$ is injective, b must go to zero under $F'\varphi$, and thus the original element was 0 in $\varinjlim F'$. □

A6.2 Flat Modules as Colimits of Free Modules

Every A-module M is the colimit of a diagram of free modules, because if

$$F \xrightarrow{\varphi} G \to M \to 0$$

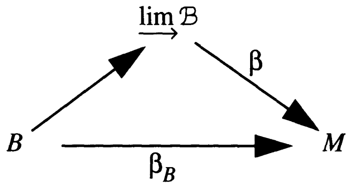
is a free presentation, then $M = \text{coker } \varphi = \text{coequalizer } (\varphi, 0)$. But there is another, less obvious, diagram of which M is the colimit. We want to define the **diagram of free modules over M** to be roughly the diagram "of free modules mapping to M and maps between them commuting with the maps to M." Because we want to take colimits only over small categories, and because we want a certain functoriality, we must exercise some care. We define a **finite list** of elements of M to be, for some natural n, a choice of n elements $m_1, \ldots, m_n \in M$. We allow repeated elements in such lists.

Definition. *The* ***diagram of free modules over*** M *is the functor* $F : \mathcal{B} \to (A\text{-Mod})$ *such that an object of* B *is a free* A*-module with a distinguished basis consisting of a finite list of elements of* M*, and whose morphisms are the maps between these free modules commuting with the natural maps of these free modules to* M*. That is, if* $\{m_i\}$ *and* $\{n_j\}$ *are finite lists of elements of* M*, and if* $B = \oplus Am_i$ *and* $C = \oplus An_j$ *are the corresponding free modules in* $\mathcal{B}$*, then a morphism of* A*-modules* $B \to C$ *is in* $\mathcal{B}$ *iff it commutes with the maps* $B \to M$ *and* $C \to M$ *sending* m_i *to* m_i *and* n_j *to* n_j*, respectively.*

Proposition A6.5. *If* M *is an* A*-module, then* M *is the colimit of the diagram* $\mathcal{B}$ *of free modules over* M*.*

Proof. For every object B of $\mathcal{B}$, we will write $\beta_B : B \to M$ for the corresponding homomorphism of free modules, and we shall write $\beta : \varinjlim \mathcal{B} \to M$ for the map induced from all the maps β_B. We will show that β is an isomorphism.

For each object B of $\mathcal{B}$ we have a commutative diagram



and since every element of M is in the image of some β_B, we see that β is surjective.

To show that β is injective, let x be an element of ker β. We must show that $x = 0$. From the construction of Theorem A6.1 we see that the direct sum of all the B_i maps onto $\varinjlim \mathcal{B}$. Thus x is in the image of a finite subsum. Since a finite direct sum of the free modules in $\mathcal{B}$ is again in $\mathcal{B}, x$ is in the image of some object B of $\mathcal{B}$.

Let $x' \in B$ be an element mapping to x in $\varinjlim \mathcal{B}$. We have $\beta_B(x') = 0$. Let B_0 be the object of $\mathcal{B}$ given by the rank-1 free A-module whose basis element is the element 0 in M. The two maps $B_0 \to B$ sending the basis element to 0 and to x' are both in $\mathcal{B}$. From the definition of $\varinjlim$, it follows that the map $B \to \varinjlim \mathcal{B}$ coequalizes these two maps, and thus sends x' to 0. This shows $x = 0$. □

Theorem A6.6 (Govorov and Lazard). *Let* M *be an* A*-module.* M *is flat iff* M *is a filtered colimit of free modules, in which case the diagram of free modules over* M *is filtered.*

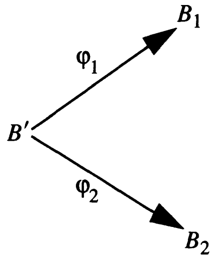
Proof. First we show that if an A-module M is a filtered colimit $\varinjlim \mathcal{B}$ of a diagram of free modules, then it is flat. If $N' \to N$ is a monomorphism of A-modules, we must prove that the induced map $M \otimes_A N' \to M \otimes_A N$ is a

monomorphism. Consider the functors F', F from $\mathcal{B}$ to (A-Mod) obtained by tensoring with N' and with N, respectively. Tensor products preserve colimits by Proposition A2.1 and Theorem A6.1 (or use Problems 2f and 3a from Appendix 5), so the map $M \otimes_A N' \to M \otimes_A N$ may be written as

$$(\varinjlim \mathcal{B}) \otimes_A N' \;=\; \varinjlim F' \;\to\; \varinjlim F \;=\; (\varinjlim \mathcal{B}) \otimes_A N.$$

However, the map $\varinjlim F' \to \varinjlim F$ is a filtered limit of monomorphisms, and is thus a monomorphism by Proposition A6.4. This shows that M is flat.

Supposing that M is flat, we must show that the diagram $\mathcal{B}$ of free modules over M is filtered. For part a of the definition it suffices to note that if B and B' are objects of $\mathcal{B}$, then $B \oplus B'$ is also an object of $\mathcal{B}$, and the inclusion maps are morphisms in $\mathcal{B}$. For part b, suppose that $\mathcal{B}$ contains maps as follows:



We must show that there are maps $\psi_i : B_i \to B$ in $\mathcal{B}$ with $\psi_1\varphi_1 = \psi_2\varphi_2$.

As a first approximation to this goal, let $C = B_1 \oplus B_2$; this is again an object of $\mathcal{B}$. Let $\beta : C \to M$ be the map sending the basis of C to the corresponding list of elements of M. By Corollary 6.6 there is a map $\gamma : C \to B$ in $\mathcal{B}$ which annihilates the kernel of β. We claim that the maps ψ_i obtained by composing γ with the inclusions $B_i \to C$ fulfill the necessary condition. Indeed, this is clear because the composite maps $B' \to B_i \to C \to B \to M$ are the same for $i = 1$ and 2, and therefore by the definition of B the composites $B' \to B_i \to C \to B$ are the same for $i = 1$ and 2.

A6.3 Colimits in the Category of Commutative Algebras

The case of modules is simple and entirely familiar—see Exercise A6.1 for a review. Here we will describe the situation in the category of algebras. See the exercises for further examples.

Proposition A6.7. *Limits and colimits exist in the category of algebras over a commutative ring A. More precisely:*

a. *Limits are set-theoretic; that is, the limit of any diagram has underlying set equal to the limit of the corresponding diagram of sets. In more detail, the product of a family of algebras B_i is the direct product of the B_i as sets, with componentwise addition and multiplication; and the equalizer of a pair of maps of algebras*

$$C_1 \underset{\psi'}{\overset{\psi}{\rightrightarrows}} C_2$$

is the subalgebra $\{c \in C_1 | \psi c = \psi' c\}$ of C_1.

b. *The coproduct of a family of algebras $\{B_i\}_{i\in\Lambda}$ is the "restricted tensor product," written $\otimes_A B_i$, whose elements are finite sums of tensors*

$$\cdots \otimes 1 \otimes 1 \otimes b_i \otimes 1 \otimes 1 \otimes \cdots \qquad b_i \in B_i$$

all but finitely many of whose factors are 1. This may also be defined to be the colimit of the category whose objects are tensor products of finitely many of the B_i, with a morphism $\otimes_{A,i\in\Lambda'} B_i \to \otimes_{A,i\in\Lambda''} B_i$ for each inclusion of finite subsets $\Lambda' \subset \Lambda''$ of Λ, defined by tensoring with the identity elements of the B_j with $j \in \Lambda'' - \Lambda'$.

c. *The coequalizer of a pair of maps of A-algebras*

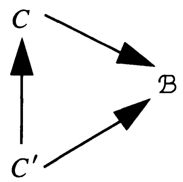
$$C_1 \underset{\psi'}{\overset{\psi}{\rightrightarrows}} C_2$$

is the algebra C_2/I, where I is the ideal generated by all elements of the form $\psi(c) - \psi'(c)$ for $c \in C_1$.

Proof.

a. Given a diagram $\mathcal{B}$ of algebras, we may regard each algebra as a set, and thus think of $\mathcal{B}$ as a diagram of the category of sets (that is, take the image of the "forgetful functor"). Let $C = \varprojlim \mathcal{B}$ in the category of sets. In the case of products and equalizers, C has a natural algebra structure, presented in the proposition. With this structure, the universal morphism $C \to \mathcal{B}$ in the category of sets, coming from the definition of the limit, is actually a map of algebras. It follows from the construction of limits dual to Theorem A6.1 that also in the general case C has a natural algebra structure, making the universal morphism $C \to \mathcal{B}$ in the category of sets into a morphism in the category of algebras.

To show that C is really the limit in the category of algebras, suppose that $C' \to \mathcal{B}$ is a morphism in the category of algebras. Again we may regard it as a morphism in the category of sets, so we see that there is a unique morphism of sets $C' \to C$ making the diagram



commutative. It suffices to show that this is actually a map of algebras, and to do this it suffices again to treat the cases of products and equalizers. In these cases the result is immediate.

b. Because of the definition, it is enough to check that the coproduct of a finite set of algebras $B_1, \ldots, B_n$ is the tensor product of those algebras over A. For simplicity, we write $\otimes$ for $\otimes_A$.

First, there are natural maps $B_i \to \otimes_i B_i$ sending b_i to the tensor product of b_i with the identity elements of the B_j for $j \neq i$.

Next, given algebra maps φ_i from each B_i to some A-algebra C', we must show that there is a unique A-algebra map $\varphi : \otimes_i B_i \to C'$ whose composition with the natural map $B_i \to \otimes_i B_i$ is φ_i. Since $\otimes_i B_i$ is generated as an algebra by the images of the individual B_i, the uniqueness statement is clear, and it suffices to prove existence.

There is a multilinear map $\Pi_i B_i \to C'$ sending

$$\sum_{j=1}^{m} (b_{1j}, \ldots, b_{nj})$$

to

$$\sum_{j=1}^{m} (b_{1j} \cdot \cdots \cdot b_{nj}),$$

where the product $\cdot$ is the multiplication of the algebra C'. By the definition of the tensor product (Appendix 2), there is a corresponding map $\otimes_i B_i \to C'$ sending

$$\sum_{j=1}^{m} (b_{1j} \otimes \cdots \otimes b_{nj})$$

to

$$\sum_{j=1}^{m} (b_{1j} \cdot \cdots \cdot b_{nj}).$$

This is evidently an algebra map, proving existence.

c. Clear from the definitions. □

A6.4 Exercises

Exercise A6.1: In the category of modules over a ring A, show that the product is the direct product, while the coproduct is the direct sum. (In the infinite case, these are not the same. The direct product of infinitely many modules M_i is the set of sequences of elements (m_i) with $m_i \in M_i$; the direct sum is the set of such sequences for which only finitely many m_i are nonzero.) Construct equalizers and coequalizers in terms of kernels and cokernels.

Exercise A6.2: In the category of sets, show that the product is the direct product, while the coproduct is the disjoint union. What are equalizers and coequalizers?

Exercise A6.3: In the category of (not necessarily Abelian) groups, show that the product is the direct product, while the coproduct is the free product. What are the equalizers and coequalizers?

Exercise A6.4: Show that any module is the filtered colimit of its finitely generated submodules, the maps being the inclusions.

Exercise A6.5: Show that any module is the filtered colimit of finitely presented modules.

Exercise A6.6: Show that any A-algebra B is the colimit of the category whose objects are the finitely generated subalgebras of B, and whose maps are the inclusions.

Exercise A6.7: Show that if B is an A-module and S is a multiplicatively closed subset of A, then the localization $B \to B[S^{-1}]$ is the colimit of the diagram of localizations $B[t^{-1}]$ for $t \in S$, where the maps are the natural maps $B[t^{-1}] \to B[(tt')^{-1}]$. If B is an A-algebra, show that the same is true in the category of A-algebras.

Exercise A6.8: State and prove the analogue of Proposition A6.3 in the category of A-algebras. Show that if $\mathcal{B}$ is a filtered subcategory of the category of A-algebras, then

a. If all the objects of $\mathcal{B}$ are domains, then $\varinjlim \mathcal{B}$ is a domain.

b. If all the objects of $\mathcal{B}$ are fields, then $\varinjlim \mathcal{B}$ is a field.

For the case of general colimits, the situation is much more delicate; see Exercise A1.1 for a special case.

Exercise A6.9: Let $\mathcal{B}'$ be a set of objects of a category $\mathcal{A}$ and a set of morphisms between these objects. Let $\mathcal{B}$ be the subcategory generated by

$\mathcal{B}'$, which may be defined either as the smallest subcategory containing $\mathcal{B}'$ or as the subcategory with the same objects as $\mathcal{B}'$ whose morphisms are all the possible compositions of morphisms of $\mathcal{B}'$ and identity morphisms. Show that $\varinjlim \mathcal{B}'$ exists iff $\varinjlim \mathcal{B}$ exists, and that if they exist they are canonically isomorphic.

Exercise A6.10 (Limits and colimits are not exact): It is shown in the text that filtered colimits are exact in the category of modules. Give an example to show that, in the category of modules over a ring, equalizers (kernels) are not right exact and coequalizers (cokernels) are not left exact; thus limits and colimits in general are not exact.

Exercise A6.11 (Filtered limits are not exact): It is shown in Chapter 7 that certain limits in the category of modules over a ring are exact. Show that in general even filtered limits in this category are not right exact by taking the limit of the following exact sequences

$$\begin{array}{lcccccccccc} E_{n+1}: & 0 & \to & \mathbf{Z} & \xrightarrow{p^{n+1}} & \mathbf{Z} & \to & \mathbf{Z}/p^{n+1} & \to & 0 \\ & & & \downarrow p & & \| & & \downarrow & & \\ E_n: & 0 & \to & \mathbf{Z} & \underset{p^n}{\to} & \mathbf{Z} & \to & \mathbf{Z}/p^n & \to & 0. \end{array}$$

Appendix 7
Where Next?

So [said the doctor]. Now vee may perhaps to begin. Yes?

—The last line, delivered by Portnoy's psychiatrist, after three hundred pages of Portnoy's confessions, in "Portnoy's Complaint," by Phillip Roth (1967).

Many references to the literature of commutative algebra are scattered through this book, in the hope that the reader will be attracted to look beyond what I have been able to include. Nonetheless it seemed worthwhile to collect here a list—even if idiosyncratic, heterogeneous, and incomplete—of readings that might make suitable "next steps" in commutative algebra and algebraic geometry. Some, like the marvelous books of Serre, are relatively self-contained. Others will require backtracking through references. I apologize in advance to all the authors whose beautiful and worthwhile books and papers I wasn't clever enough to include! The order is roughly from more to less geometric.

Basic Algebraic Geometry: Here is a path I might suggest for my own students: Read parts of the books of Harris [1992], Mumford [1976], Eisenbud and Harris [1992]; then Hartshorne [1977] balanced by Shafarevich [1972] and Griffiths and Harris [1978]. Be sure and go on to deeper things before spending too much time on the basics!

Some further steps in Algebraic geometry: Mumford [1966], and [1975], Beauville [1983], Arbarello, Cornalba, Griffiths, and Harris [1985].

Canonical curves: For the geometric side of the complexes in Appendix 2: St. Donat [1973]. The paper of Schreyer [1991] relates canonical curves to Gröbner bases.

Elliptic curves are a topic on the border of number theory and algebra geometry. An excellent introduction is Silverman [1986]. For a look at the state of the art, see Cornell and Silverman [1986].

Rings of Witt vectors and the number-theoretic aspect of complete local rings: Serre [1979]. This book also has a nice treatment of Galois descent, carried further in Serre [1973].

Spectral sequences and other homological algebra, with geometric applications in mind: Grothendieck [1957].

The fundamental group and étale cohomology on the geometric side, and the theory of Hensel rings and étale extensions on the algebraic side: Grothendieck [1971], Artin [1973], and Milne [1980] (parts of which should probably be read first). Add Azumaya [1950] for the truth about lifting idempotents.

Intersection Theory: Serre [1957], Fulton [1983] and [1984], and for the arithmetic connection, Szpiro [1987]. For a beautiful application of Fulton's Theory to commutative algebra see the paper by Roberts in Hochster, Huneke, and Sally [1989].

Determinantal rings and varieties: Bruns and Vetter [1988], DeConcini, Eisenbud and Procesi [1980, 1982]. For a surprising application to geometry see Gruson, Lazarsfeld, and Peskine [1983]. A current development of interest is given in Conca and Herzog [in press]. For a look at the (possible) future, see Gelfand, Kapranov and Zelevinsky [1993].

Primary decomposition and computation: Gianni, Trager, and Zacharias [1989] and Eisenbud, Huneke and Vasconcelos [1992] for two views of the problem of computing primary decomposition. Eisenbud and Sturmfels [1994] for a recent piece of theoretical work with computational overtones.

Recent work in commutative algebra: The book on Cohen-Macaulay rings by Bruns and Herzog [1993] is an easy way to get a look at what has been going on lately. A broad view of the state of the field in 1987 may be had from the volume of papers edited by Hochster, Huneke, and Sally [1989].

Recent progress in methods of commutative algebra using positive characteristic: The papers of Hochster, Huneke, Watanabe, and others, starting with Hochster and Huneke [1990]. Much of this work is related to a group of homological conjectures nicely presented in Hochster [1975].

Infinite free resolutions: Gulliksen and Levin [1969]. For a recent development see Avramov, Gasharov, Peeva [1994].

Hints and Solutions for Selected Exercises

Chapter 1

Exercise 1.1:

$1 \Rightarrow 2$: If $I_1 \subsetneqq I_2 \subsetneqq \cdots \subset M$, let $I = \cup_j I_j$. If I is generated by $f_1, \ldots, f_n$, then each f_i must be in one of the I_j, so from some i on, $I_i = I$.

$2 \Rightarrow 3$: Given any set of submodules, choose a maximal ascending sequence in it—this must be finite by condition 2. (Note that we don't even need Zorn's lemma for this.)

$3 \Rightarrow 4$: If 4 did not hold, the sequence of submodules $I_j = (f_1, \ldots, f_j)$ would have an infinite strictly ascending subsequence.

$4 \Rightarrow 1$: If a submodule $I \subset M$ I were not finitely generated, then it would be possible to choose $f_1, f_2, \ldots \in I$, contradicting condition 4, by choosing f_{i+1} in I but not in the submodule generated by $f_1, \ldots, f_i$.

Exercise 1.4:

$1 \Rightarrow 2$: If R is Noetherian then R_0 is Noetherian because it is a homomorphic image of R. The second condition is trivial because every ideal of R is finitely generated.

$2 \Rightarrow 3$: Assume that $x_1, \ldots, x_r$ generate the ideal $R_+ = R_1 \oplus R_2 \oplus \ldots$. Replacing the x_i by their homogeneous parts, we may assume that

they are homogeneous. We claim that they generate R as an R_0-algebra. It suffices to show that each homogeneous element $f \in R$ is in the subalgebra R' generated by the x_i. Since f is homogeneous, we may write $f = \sum_i a_i x_i$ with homogeneous coefficients a_i (given a representation with non-homogeneous a_i, we could replace each a_i by its homogeneous component of degree $\deg f - \deg x_i$). Since the x_i have strictly positive degrees, each a_i may be chosen to have degree less than that of f. By induction on the degree, the a_i belong to R', and the given expression for f then shows that f does too.

$3 \Rightarrow 1$: Apply Corollary 1.3. □

Exercise 1.5: If $I_1 \subset I_2 \subset \cdots$ is an ascending chain of ideals of R, then since S is Noetherian we must have

$$I_n S = I_{n+1} S = \cdots$$

for some n. But $\pi(I_j S) = I_j \pi(S) = I_j R = I_j$ for any j, so

$$I_n = I_{n+1} = \cdots . \square$$

Exercise 1.6:

b. Along with any monomial $A = x_1^{m_1} \cdots x_r^{m_r}$, p must contain all the monomials $\sigma^{-1}(A) = (x_{\sigma(1)})^{m_1} \cdots (x_{\sigma(r)})^{m_r}$ where $\sigma \in \sum$ is a permutation.

e. To prove that any $f \in S^{\Sigma}$ can be written as a polynomial in the f_i, consider the initial term of f. Using conditions b and c we see that some monomial in the f_i has the same initial term. Subtracting this monomial, we get an invariant with a smaller initial term. We repeat the process; by condition a, it cannot go on forever.

   To prove the uniqueness of the expression, it is enough to show that for any nonzero polynomial $q(y_1, \ldots, y_r)$, the polynomial $p(x_1, \ldots, x_r) = q(f_1(x_1, \ldots, x_r), \ldots, f_r(x_1, \ldots, x_r))$ is nonzero. By condition d, the initial terms of distinct monomials in the f_i are distinct monomials in the x_i. Thus there is a unique term $a y_1^{\mu_1} \cdots y_r^{\mu_r}$ of $q(y_1, \ldots, y_r)$, with $a \in k$, for which the monomial in the initial term of $a f_1(x_1, \ldots, x_r)^{\mu_1} \cdots f_r(x_1, \ldots, x_r)^{\mu_r}$ is maximal. This uniqueness implies that the initial term of $p(x_1, \ldots, x_r)$ will be the initial term of $a f_1(x_1, \ldots, x_r)^{\mu_1} \cdots f_r(x_1, \ldots, x_r)^{\mu_r}$. In particular, $p(x_1, \ldots, x_r) \neq 0$.

Exercise 1.7:

a. To show that $k[x^2, xy, y^2]$ is not isomorphic to a polynomial ring, note that it does not have unique factorization ($x^2 \cdot y^2 = xy \cdot xy$).

b. For $m \in \mathbf{Z}^r$, let $\alpha(m) \in \mathrm{Hom}(G, k^*)$ be the map $g \mapsto g(x^m)/x^m$. The invariants are generated by the monomials whose exponents lie in ker α. Use the fact that k is a field (and thus k^* cannot have more than d elements of order d for any integer d) to show that the group $\mathrm{Hom}(G, k^*)$ is finite. Thus ker α is itself isomorphic to $\mathbf{Z}^r$. If $t_1, \ldots, t_r$ is a basis for ker α, show that the ring of invariants has quotient field $k(t_1, \ldots, t_r)$.

Exercise 1.11: The map $t \mapsto (4t/(t^2+4), (t^2-4)/(t^2+4))$, which is a stereographic projection from the "north pole," gives a bijection $\mathbf{C} - \{\pm 2i\}$ to $\{(x,y) | x^2+y^2=1\} - \{(0,1)\}$. (To improve this bijection, try moving $2i$ "to ∞" with the substitution $t = \frac{1}{s} + 2i$.)

Exercise 1.14: Show that the vector space $k[x,y]/(y^2-x^3)$ is spanned by the classes of elements of the form x^n and $x^n y$ with $0 \le n \le \infty$, and show that the images of these elements in $k[t]$ form a basis for $k[t^2,t^3] = k[t^2,t^3,t^4,\ldots]$.

Exercise 1.15b: If k is a field of characteristic not 2 and Q is a quadratic form in n variables, regarded as a function of a vector $v \in k^n$, then show that the function of two vector variables $\langle v, w\rangle = (Q(v+w) - Q(v) - Q(w))/2$ is a symmetric bilinear form on k^n, and $Q(v) = \langle v, v\rangle$. Thus the classification of quadratic forms is the same as the classification of symmetric bilinear forms, which is the same as the classification of symmetric $n \times n$ matrices up to the equivalence relation $M \sim AMA^{\text{transpose}}$. We define the **rank** of the form to be the rank of the corresponding matrix M.

Show that given any symmetric bilinear form on k^n, there is a basis of k^n consisting of vectors orthogonal to one another; that is, any symmetric matrix is equivalent to a diagonal matrix. If k is algebraically closed, then in addition the diagonal entries can be made to be ones and zeros, so the forms are classified by the number of ones on the diagonal, the rank.

Exercise 1.16b: Prove that if a polynomial over an infinite field vanishes on a line through the origin, then so does each of its homogeneous components.

Exercise 1.17: Consider the homogenization of the element $(x_1^2+x_2) - (x_1^2+x_3) = x_2 - x_3$.

Exercise 1.20: Try an ideal in $k[x_1,x_2]$ generated by two forms of the same degree having relations only in a high degree like $(x_1^d x_2^e, x_1^e x_2^d)$ with $d \ll e$.

Exercise 1.21a: $F(n) = F(0) + \sum_0^{n-1} G(n)$, so it is enough to show that such sums of polynomials are polynomials of degree one more. Since the

polynomial F_k of part b has degree exactly k, any rational polynomial can be expressed as a rational linear combination of the F_k, and it is enough to check the desired property for F_k. In fact we have

$$\sum_{m=0}^{n-1} F_k(m) = \sum_{m=0}^{n-1} \binom{m}{k} = \binom{n}{k+1}, = F_{k+1}(n).$$

This holds because a choice of $k+1$ elements from $\{1,\ldots,n\}$ may be thought of as a choice of a last element, call it $m+1$, together with a choice of k elements from $\{1,\ldots,m\}$.

Chapter 2

Exercise 2.11: $u(n/u) = n/1$.

Exercise 2.19: a. If m goes to zero in each $M[f_i^{-1}]$, then m is killed by a power of each f_i. Show that if the set $\{f_i\}$ generates the unit ideal, then so does the set $\{f_i^{n_i}\}$ for any positive integers n_i.

b. Since m_i and m_j become equal in $M[f_i^{-1}f_j^{-1}]$, we must have $(f_if_j)^N(m_i - m_j) = 0$ for large N. We may as well assume that the set of f_i is finite (if $1 \in (\{f_i\})$, then it is in the ideal generated by finitely many of the f_i) so some power N will do for all i, j.

Using the idea of part a, we may write

$$1 = \sum r_i f_i^N$$

(this is the "partition of unity"). We may suppose that N is large enough so that for each i the element $f_i^N m_i$ is the image of some element $m_i' \in M$. Show that $m = \sum_i r_i m_i'$ has the desired property.

Exercise 2.22: If I is the ideal and is not prime, then there exist $a, b \in R$, not in I, with $ab \in I$. Show that (I, a) can be generated by a and finitely many elements g_j of I. Then show that

$$J := (I : a) = \{r \in R | ra \in I\}$$

is finitely generated, and that aJ together with the g_j generate I, contradicting the hypothesis.

Exercise 2.27:

a. After proving that $M = \Pi_{\gamma\in\Gamma} M_\gamma$, prove the fact about homomorphisms with the same idea: $\varphi_\gamma = e_\gamma\varphi$. The uniqueness of the decomposition follows, since φ is an isomorphism iff all the φ_γ are isomorphisms.

b. Consider the direct sum $R = \oplus_\gamma R_\gamma$.

Chapter 3

Exercise 3.3: Since all the sets of primes involved behave well with respect to localization, it suffices to prove that if R is a local ring with maximal ideal P then P is in Ass $\mathrm{Hom}_R(M, N)$ iff it is in $\mathrm{Supp}\, M \cap \mathrm{Ass}\, N$. If P is in $\mathrm{Supp}\, M$, show using Nakayama's lemma that there is a surjection $M \twoheadrightarrow R/P$. If P is also in $\mathrm{Ass}\, N$, there is an inclusion $R/P \subset N$. The composition $\varphi \in \mathrm{Hom}_R(M, N)$ of these two maps is annihilated by P, so $P \in \mathrm{Ass}\,\mathrm{Hom}_R(M, N)$. Conversely, if $P \in \mathrm{Ass}\,\mathrm{Hom}_R(M, N)$, then we can choose $0 \neq \varphi \in \mathrm{Hom}_R(M, N)$ with annihilator P. It follows that $M \neq 0$, so $P \in \mathrm{Supp}\, M$, and also that $\mathrm{im}\, \varphi \subset N$ is annihilated by P, so $P \in \mathrm{Ass}\, N$.

Exercise 3.4:

a. The inclusion $\mathrm{Content}(fg) \subset \mathrm{Content}(f)\mathrm{Content}(g)$ is obvious. To prove the second inclusion, it is enough to show that if a prime $P \subset R$ contains $\mathrm{Content}(fg)$, then it contains $\mathrm{Content}(f)\mathrm{Content}(g)$. Factoring out P, we may assume that R is a domain and P is 0, and we must show that if $fg = 0$ then f or $g = 0$. Since S is now a domain, this is obvious.

b. If R is Noetherian and $\mathrm{Content}(f)$ consists of zerodivisors, then $\mathrm{Content}(f)$ annihilates a nonzero element of R by Corollary 3.2. It follows that f annihilates this same element viewed as an element of S.

c. Use part a to show that for any polynomials f, g, and $h \in R[x]$:

   *) If $rf = gh$ in $R[x]$ with $r \in R$, and p is a prime of R dividing r, then p divides g or h.

   Use * to show that if f is irreducible in $R[x]$ and R is factorial then f is irreducible in $K[x]$.

   Now suppose that R is factorial. To show that $R[x]$ is factorial, verify that $R[x]$ satisfies ascending chain condition on principal ideals and that an irreducible polynomial $f \in R[x]$ is prime. Using the fact that f is prime in $K[x]$, this reduces to showing that if $fg = rh$ in $R[x]$ with $r \in R$, then r divides g. Use * again for this.

Exercise 3.6: Ideals generated by subsets of the variables are prime. Ideals generated by powers of some of the variables are irreducible. Ideals generated by square-free (that is, multilinear) monomials are radical. Ideals containing a power of each of a certain subset of the variables, and generated by elements involving no further variables, are primary.

Exercises 3.7 and 3.8: Let I be a monomial ideal. The key point is that if m is a minimal generator of I, and we can factor m into relatively prime

parts $m = m'm''$, then $I = (I + (m')) \cap (I + (m''))$. It is also useful to note that a monomial n is in I iff it is divisible by one of the minimal generators of I.

Exercise 3.9:

a. Do induction on r, inverting x_r and using Theorem 3.1.

b. With an induction as in the hint for part a, it suffices to show that $\mathfrak{m} = (x_0, \ldots, x_r)$ is an associated prime iff Γ is connected and all the variables are involved in the I_j. If Γ is connected, let T be a spanning tree (that is, a connected subgraph of Γ containing all the vertices, and whose edges form no loops). For each edge e of T, let i_e be an element of the intersection of the two sets corresponding to the vertices incident to e. Show that $\mathfrak{m}$ is associated by showing that $\mathfrak{m}\,\Pi_{e \in T} x_{i_e} \subset I$. Conversely, if Γ is not connected, we may partition Γ into two subgraphs Γ_1 and Γ_2 that are not connected to each other. Let I_1 and I_2 be the corresponding subproducts of primes, so that $I = I_1 I_2$, and I_1 and I_2 involve disjoint subsets of the variables. Show that $I_1 I_2 = I_1 \cap I_2$, and thus the associated primes of R/I are those of I_1 (which involve only the first set of variables) and those of I_2, involving only the second.

Exercise 3.10: a. As a vector space the ring R is $k[b] \oplus ka$.

d. J_1 must contain all forms of degree greater than or equal to d, where d is the maximal degree of a nonzero homogeneous element of $(0 : (x_1, \ldots, x_r))$.

Exercise 3.11: First, let $\mathcal{S} = x_1, \ldots, x_s$ be a subset of the variables $x_1, \ldots, x_r$. Relating each monomial ideal to its minimal set of monomial generators, observe that there is a one-to-one correspondence between monomial ideals primary to $(\mathcal{S})$ and monomial ideals primary to the irrelevant ideal in $K[x_1, \ldots, x_s]$, where $K = k(x_{s+1}, \ldots, x_r)$ is the field of rational functions in the remaining variables.

Use this observation, together with the fact that primary decompositions localize, to reduce the problem to the following special case:

Let $\mathfrak{m}$ be the irrelevant ideal of $S = k[x_1, \ldots, x_r]$. Suppose that I is a monomial ideal of S, and $I = I' \cap J$ where I' is a monomial ideal having no $\mathfrak{m}$-primary component and J is an $\mathfrak{m}$-primary monomial ideal. In this case the unique maximal monomial choice of J is the ideal generated by all those monomials *not* dividing any of the finitely many monomials in I' but not in I. (The finiteness of this set implies that J contains a power of each variable and is thus $\mathfrak{m}$-primary.)

Exercise 3.15: a. In this case all the prime ideals of $K(R)$ are maximal ideals of $K(R)$.

Exercise 3.18: Suppose that $f \in \cup_{i \in B} P_i$, and suppose that

$$\{P_1, \ldots, P_n\} = \{P_i | \text{ some monomial of } f \text{ is in } P_i\}.$$

Show that if g is a polynomial such that g and $f+g$ are in $\cup_{i \in B} P_i$, then $g \in \cup_{m=1}^{n} P_m$. Conclude that if $f \in I \subset \cup_{i \in B} P_i$, then $I \subset \cup_{m=1}^{n} P_m$ and I is thus contained in one of the $P_1, \ldots, P_n$.

Exercise 3.19:

a. Regard the subspace in R generated by $f_1, \ldots, f_n$ as an image of an n-dimensional vector space V. Let W_j be the preimage of I_j in V, and write $f_1, \ldots, f_n$ again for the basis elements of V. Let m_j be the dimension of W_j. Consider the $n \times (m_j + 1)$ matrix M_j whose first column is $(t_1, \ldots, t_n)$ and whose other m_j columns represent a basis for W_j. The $(m_j + 1) \times (m_j + 1)$ minors of M_j are linear forms in the t_i. For each j, the condition $\sum a_i f_i \notin W_j$ is the condition that one of these forms is nonzero at $(a_1, \ldots, a_n)$. Since not all of the f_i are in W_j, one of these linear forms is not identically 0; call it L_j. The polynomial g may be taken to be $\Pi_j L_j$.

b. We do induction on n, the case $n = 1$ being trivial. We may suppose that none of the primes $I_1, \ldots, I_n$ contains another of the I_j. By induction we may choose $j_1 \in J$ such that $(f + j_1) \notin \cup_{k=2}^{n} I_k$. If $(f + j_1) \notin I_1$ we are done, so suppose that $(f + j_1) \in I_1$.

   Since $f + j_1 + J = (f + J) \not\subset I_1$, we must have $J \not\subset I_1$. Since I_1 is prime we therefore have $J \frown \cap_{k=2}^{n} I_k \not\subset I_1$. Let $j_2 \in J \frown \cap_{k=2}^{n} I_k$ be an element outside I_1. It is easy to see that $f + j_1 + j_2 \notin \cup_{k=1}^{n} I_k$.

Chapter 4

Exercise 4.4: $tx^3 + tx^2 - x^2 - x = (tx - 1)(x^2 + x)$, and $tx - 1$ is a unit.

Exercise 4.11: a. Choose a minimal set of generators $g_1, \ldots, g_n$ for M. By Nakayama's lemma, the images of the g_i form a basis of the R/P-vector space M/PM.

Let $F = R^n$ be a free module of rank n, and let $\varphi : F \to M$ be the map sending the ith generator of F to g_i. Since φ induces an isomorphism $F/PF \to M/PM$, we see that $\ker \varphi \subset PF$. But φ is an epimorphism and M is projective, so φ is split, and we can write $F = M' \oplus \ker \varphi$, with $M' \cong M$. Thus $PF = PM' \oplus P \ker \varphi$. Since $\ker \varphi \subset PF$, Nakayama's lemma shows that $\ker \varphi = 0$, so $F = M', \varphi$ is an isomorphism, and M is free as required.

If R is graded, with R_0 a field, and M is a graded module, we let P be the irrelevant ideal R_+, and choose the g_i to be homogeneous elements.

Taking the ith generator of F to have degree=deg g_i, we may proceed as before.

b. It is clear from part a that if M is projective then each M_P is free over R_P; that is, M is locally free. Conversely, if M is locally free, and $\varphi : F \to M$ is a surjection, use Proposition 2.10 to show that $\mathrm{Hom}_R(M, F) \to \mathrm{Hom}_R(M, M)$ is a surjection. Any map $\psi \in \mathrm{Hom}_R(M, F)$ in the preimage of 1 is a splitting of φ.

Exercise 4.12a: For the "intermediate step," if $\varphi : M_P \to N_P$, then by Proposition 2.10 there is map $\varphi' : M \to N$ such that $\varphi = \varphi'/f_1$ for some $f_1 \notin P$. Since φ' becomes an isomorphism on localizing, coker φ' is annihilated by some $f_2 \notin P$. We may repeat the argument with $\psi = \varphi^{-1}$, obtaining a map ψ' and elements $f_3, f_4 \notin P$. Take $f = f_1 f_2 f_3 f_4$. The maps $\varphi'\psi'$ and $\psi'\varphi'$ are both epimorphisms over $R[f^{-1}]$, so they are both isomorphisms by Corollary 4.4a.

Now suppose that M is projective. By Exercise 4.11b it becomes free when we localize at any prime. The "intermediate step" provides an element $f_P \notin P$ such that $M[f^{-1}]$ is free. Clearly, $(\{f_P | P \text{ a prime}\}) = R$, so some finite set of f_P already generates the unit ideal.

For the converse, use the characterization of projectives as those modules M such that for some epimorphism $F \to M$ from a free module the induced map $\mathrm{Hom}_R(M, F) \to \mathrm{Hom}_R(M, M)$ is surjective (Proposition A3.1c, together with Propositions 2.9 and 2.10; see also Exercise 2.20).

Exercise 4.13: Suppose that $P_1, \ldots, P_n$ are the maximal ideals of R. Let $\varphi_i : M_{P_i} \to N_{P_i}$ be isomorphisms of R_{P_i}-modules. Since $\mathrm{Hom}_{R_{P_i}}(M_{P_i}, N_{P_i}) = \mathrm{Hom}_R(M, N)_{P_i}$, by Proposition 2.10, we may multiply each φ_i by a unit of R_{P_i} and assume that it is the localization of a map $\psi_i : M \to N$.

Since $P_i \not\supset \cap_{j \neq i} P_j$ for each i, we may (by Lemma 3.3) choose elements $a_i \in R_i$ such that $a_i \in \cap_{j \neq i} P_j$ but $a_i \notin P_i$. We claim that $\psi := \sum_i a_i \psi_i : M \to N$ is an isomorphism.

To see this, it suffices (from Corollary 2.9) to show that ψ_{P_i} is an isomorphism for each i. In fact, we claim that if (R, P) is any local ring and $\varphi, \gamma : M \to N$ are homomorphisms between finitely generated R-modules such that φ is an isomorphism and $\gamma(M) \subset PN$, then $\varphi + \gamma$ is an isomorphism. Applying this result to the localizations at P_i of $\varphi = a_i \varphi_i$ and $\gamma = \sum_{j \neq i} a_j \varphi_j$ gives the desired conclusion.

To prove the claim, note that γ induces the map 0 from M to N/PN. Since φ induces an epimorphism $M \to N/PN$, so does $\varphi + \gamma$. By Nakayama's lemma, $\varphi + \gamma : M \to N$ is an epimorphism, and we must show that it is a monomorphism. Composing $\varphi + \gamma$ with an isomorphism $\alpha : N \to M$, we get an endomorphism of M that is surjective. By Corollary 4.4, $\alpha(\varphi + \gamma)$ is an isomorphism, and thus $\varphi + \gamma$ is a monomorphism as required. $\square$

Exercise 4.17: It is easy to show that if $R[x]$ is integrally closed in $S[x]$, then R is integrally closed in S. For the converse:

First reduce to the case where R is Noetherian by passing to a subring finitely generated over $\mathbf{Z}$. If $f(x) \in S[x]$ is integral over $R[x]$, then $M := R[x][f(x)] \subset S[x]$ is a finitely generated $R[x]$-module. Let coef (M) be the submodule of S generated by all coefficients of elements of M. Show that coef (M) is a finitely generated R-module. If α is the leading coefficient of f, show that $R[\alpha] \subset \operatorname{coef}(M)$; it follows that $R[\alpha]$ is a finitely generated R-module, so α is integral over R (Corollary 4.6). Use induction on the degree and the fact that the integral elements form a ring to show that $f \in R[x]$.

Exercise 4.18: Use Exercise 4.17, letting S be the quotient field of R. Note that $S[x]$ is factorial, and thus normal.

Exercise 4.21a: Use the characterization that an element $t \in T$ is integral iff t stabilizes some finitely generated S-submodule M of T. Show that if t stabilizes M, then the leading form of t stabilizes the module of leading forms of M. Show that if M is finitely generated, then the module of leading forms of M is finitely generated by showing that the latter is contained in the module generated by all the components of a set of generators of M.

Chapter 5

Exercise 5.7: Use the function that is $e^{\frac{-1}{x^2}}$ for $x > 0$ and 0 for $x < 0$, and its reflection in the y-axis.

Exercise 5.8b: Show first that $\operatorname{in}(I) = (\operatorname{in}(t^4), \operatorname{in}(t^5), \operatorname{in}(t^{11}))$, so it is enough to check that these three elements annihilate $\operatorname{in}(t^{11})$.

Chapter 6

Exercise 6.1b: Show that if $0 \to M' \to F \to M \to 0$ is a short exact sequence with M and F flat, then M' is flat.

Exercise 6.2a: Assuming that R is local, choose a minimal set of generators for M, and use these to define a map from a free module F onto M; let K be the kernel. Show that the construction can be done so that K is finitely generated. Now tensor with R/P, and use Nakayama's lemma to conclude that $K = 0$ iff $\operatorname{Tor}_1^R(M, R/P) = 0$ iff M is flat. This argument also gives an alternate (and much simpler) proof of Theorem 6.8 in the case where $R = S$.

Exercise 6.3: To show in the second case that $T := \mathrm{Tor}_1^R(R/(x,y), M) = 0$, note that multiplication by x is an isomorphism on M, so it acts as an isomorphism on T. However, note that it acts as 0 on $R/(x,y)$, so it acts as 0 on T.

Exercise 6.4: The short exact sequence $0 \to S \to S \to S/(f) \to 0$ gives rise to a long exact sequence in Tor. Using this with the fact that S is a free R-module, we see that $\mathrm{Tor}_1^R(R/I, S/(f)) = (0 :_{S/IS} f)$, the kernel of multiplication by f on S/IS. If I contains the coefficients of f, then this multiplication is 0, so Tor is nonzero. If the coefficients of f generate the unit ideal, then f is a nonzerodivisor on S/IS for every ideal I of R (use Gauss' lemma, Exercise 3.4).

Exercise 6.6a: Use the definition of flatness and the associativity of the tensor product,

$$(S \otimes_R T) \otimes_T M = S \otimes_R M,$$

if M is a T-module.

Exercise 6.10:

a. Use the equational criterion: The relation $\sum n_i m_i$ may be broken into homogeneous parts, and then one can replace each of the other quantities that appear by its homogeneous part as well.

b. By part a, it is enough to show that if I is homogeneous and $\mathrm{Tor}_1^R(R/I, M) \neq 0$, then its localization at $P \neq 0$. This is true for any graded module.

Exercise 6.11c: Note that $M[f^{-1}]_0$ is the direct limit of the diagram of R_0-modules

$$\cdots \xrightarrow{f} M_i \xrightarrow{f} M_{i+1} \xrightarrow{f} \cdots .$$

This makes it easy to show that if M_i is free over R_0 for all $i \gg 0$, then $M[f^{-1}]_0$ is flat over R_0. (The same technique is used in the easy half of the proof of the Govorov-Lazard theorem, Theorem A6.6.)

For the converse, show first that if $f \in R_1$, then $M[f^{-1}] = R_0[x, x^{-1}] \otimes_{R_0} M[f^{-1}]_0$. (See Exercise 2.17 for a similar situation.) Thus $M[f^{-1}]_0$ is flat over R_0 iff $M[f^{-1}]$ is flat over R_0. Now if $M[f^{-1}]$ is flat over R_0, and $I \subset R_0$ is the maximal ideal, then the kernel K of the multiplication map $I \otimes_{R_0} M \to M$, which is a finitely generated graded R-module, satisfies $K[f^{-1}] = 0$. If this is true for every $f \in R_1$, show $K_i = 0$ for $i \gg 0$. Since $I \otimes_{R_0} M \cong \oplus(I \otimes_{R_0} M_i)$ and $M \cong \oplus M_i$, we see that the multiplication maps $I \otimes_{R_0} M_i \to M_i$ are injective for all $i \gg 0$. This shows that M_i is flat over R_0, and since it is also finitely generated and R_0 is local, M_i is free.

Chapter 7

Exercise 7.6: $\mathfrak{m}$ is a union of the cosets of $\mathfrak{m}_j$ it contains.

Exercise 7.9: To exhibit the isomorphism, let (r_i) be a Cauchy sequence, and using the definition of Cauchy sequence, choose an increasing function $\sigma : \mathbf{N} \to \mathbf{N}$ such that $r_{\sigma(i)} - r_j \in \mathfrak{m}_i$ for all i and all $j > \sigma(i)$. The isomorphism is then given by sending (r_i) to the element $(r_{\sigma(i)} + \mathfrak{m}_i)$, which is independent of the choice of σ. The inverse map may be taken to be

$$(r_i + \mathfrak{m}_i) \mapsto (r_i).$$

The ambiguity introduced by the choice of $r_{\sigma(i)} \in (r_{\sigma(i)} + \mathfrak{m}_i)$ is swallowed by the equivalence relation on the Cauchy sequences.

Exercise 7.11: Regard R as an $R[x_1, \ldots, x_n]$-module by means of the homomorphism $\varphi : R[x_1, \ldots, x_n] \to R$ sending x_i to a_i, and complete the exact sequence of $R[x_1, \ldots, x_n]$-modules

$$0 \to (x_1 - a_1, \ldots, x_n - a_n) \to R[x_1, \ldots, x_n] \xrightarrow{\varphi} R \to 0$$

with respect to the ideal $(x_1, \ldots, x_n)$.

Exercise 7.14: Use induction on n. Then show that $R[[x]]$ is flat over R. Now use Exercise 7.13 to reduce to the statement of Exercise 7.12.

Exercise 7.19:

a. $A = S[x]/(g_1)$ is a free S-module on the elements $1, x, x^2, \ldots, x^{d-1}$. Uniqueness comes because g_1, being monic, is a nonzerodivisor of $S[x]$.

b. To show that G_1, G_2 generate the unit ideal of $R[x]$ is equivalent to showing that G_2 generates the R-module $M = R[x]/(G_1)$. Because G_1 is monic, M is a finitely generated R-module. Now use Nakayama's lemma.

Exercise 7.20: Reducing mod $\mathfrak{m}$ and writing h_i for $\bar{H}_i$ produces

$$0 = g_1 h_1 + g_2 h_2 \quad \text{with } h_i \in (R/\mathfrak{m})[x], \deg h_2 < \deg g_1.$$

Now use the uniqueness in part a of the lemma.

Show that the elements

$$f_1 = G_1 + H_2$$
$$f_2 = G_2 + H_1$$

satisfy the theorem.

To do the general case, find successive liftings of the factorization mod $\mathfrak{m}$, mod $\mathfrak{m}^2$, mod $\mathfrak{m}^4, \ldots$, and take the limit.

Exercise 7.23: $p(x) = 3x^2 - 2x^3$.

Exercise 7.27: The form of the criterion will be as follows: u is congruent to an nth power mod m, for an integer m that you should determine.

Chapter 9

Exercise 9.6: Let $I \subset S$ be an ideal and $s \in I$ a nonzero element. Use Corollary 2.9 to show that s, together with a set of elements that generate I locally at every maximal ideal containing s, generates I.

Chapter 10

Exercise 10.2: By localization, we may reduce to the case where R is local with maximal ideal P. To show that the primes described are the only ones contracting to P, factor out P and reduce to the case where R is a field.

a. To show that $\operatorname{codim} PR[x] \leq c$: By the converse to the PIT there are elements $a_1, \ldots, a_c \in P$ such that P is nilpotent mod $(a_1, \ldots, a_c)$. It follows that $PR[x]$ is nilpotent mod $(a_1, \ldots, a_c)R[x]$, and the PIT gives the desired conclusion.

b. Let $Q = \{g \in R[x] |$ for some $a \in R - P, ag \in PR[x] + (f)\}$. To show that $\operatorname{codim} Q \leq c + 1$, take $(a_1, \ldots, a_c)$ as above, and show that Q is minimal over $(a_1, \ldots, a_c, f)R[x]$.

Exercise 10.4: Compare $R[x]/(ax - b)$ with the domain obtained from R by adjoining the fraction b/a.

Exercise 10.9a: Assuming that the 1, 1 entry of M is the unit, we may use row and column operations to reduce to the case where the other entries of the first row and column are 0. These operations do not change $I_k(M)$.

Exercise 10.10: It suffices by Exercise 10.9 to prove that $\operatorname{codim} I_k(M) \geq (p - k + 1)(q - k + 1)$. Do induction on k, the case $k = 1$ being easy. Let P be a prime ideal containing $I_k(M)$. Suppose that $\operatorname{codim} P < (p - k + 1)(q - k + 1)$. Observe that P cannot contain all the variables. Thus we may begin by inverting a variable, which we may suppose to be x_{11}. As in Exercise 10.9, show that this reduces the problem to computing the codimension of $I_{k-1}(M')$, where M' is a smaller matrix. The entries of M' have the form $y_{ij} = x_{ij} - x_{i1}x_{1j}x_{11}^{-1}$. Show that $k[x_{ij}] = k[\{y_{ij}\}, \{x_{i1}\}, \{x_{1j}\}]$; that is, the y_{ij} may be taken as a subset of the indeterminates. Now use the inductive hypothesis on $I_{k-1}(M')$.

Chapter 11

Exercise 11.1: It is easy to show that valuation rings satisfy the condition. For the converse, prove first that the fractional ideals of R form a group, totally ordered by inclusion. This group, with the opposite order, may be taken to be the value group, with valuation sending $a \in K(R)^\times$ to the fractional ideal Ra.

Exercise 11.2a: First reduce to the case $R = R_P$. If $x \in K(R)$, suppose neither x nor x^{-1} were contained in R. By the maximality of R we have $PR[x] = R[x]$, so we may write

$$1 = r_0 + r_1 x + \cdots + r_n x^n \quad \text{with } r_i \in PR.$$

Multiplying by x^{-n}, we see that x^{-1} is integral over R, and thus in R.

Exercise 11.9: We may assume that X is affine, with coordinate ring R. Any rational map $f : X \to \mathbf{P}^r$ may be given in the form

$$f : x \mapsto (\varphi_0(x), \ldots, \varphi_r(x)),$$

where the φ_i are rational functions on X. The set on which f is a morphism defined is an open set, so that it is enough to show that f is defined at some points of any codimension-1 subvariety. Let $Y \subset X$ be a subvariety of codimension 1, corresponding to a prime P of R. Since R is normal, R_P is a discrete valuation ring. Use this fact to write

$$(\varphi_0(x), \ldots, \varphi_r(x)) = (q\psi_0(x), \ldots, q\psi_r(x))$$

with $\psi_i \in R$, not all in P, and q a rational function. Deduce that f is defined wherever not all of the ψ_i are 0—in particular, f is defined at the generic point of Y.

Exercise 11.10: First suppose R is reduced. All associated primes of R are then minimal—that is, they have codimension 0. If we localize at such a prime we get a reduced zero-dimensional ring—that is, a field. This is in particular a regular ring.

Conversely, suppose that R satisfies R0 and S1. By S1, all the primes in a primary decomposition of zero must be minimal primes. Localizing at one of these primes, we get a zero-dimensional regular ring by R0. But by definition, the maximal ideal of such a ring is generated by zero elements; that is, the ring is a field. It follows that the primary components are all prime. Thus 0 is an intersection of primes in R, as required.

Exercise 11.12: Suppose $a \in R$ is a nonzerodivisor. If P is a codimension-1 prime of $(a) \subset P$, show that

$$\operatorname{length}(R_P/((a)) = \sum_{Q \subset Pa \text{ minimal prime}} e_Q \operatorname{length} R/((a) + Q),$$

where $e_Q = \text{length}(R_Q)$, so that every principal divisor is a linear combination of relations on $A_1(R)$.

Chapter 12

Exercise 12.2: Show that R may be identified with a localization of the ring $k[s^3, s^2t, st^2, t^3]$, and pull back the ideal $\mathfrak{q}$ to this ring.

Exercise 12.3: Define a surjection

$$M/\mathfrak{q}M \otimes_{R/\mathfrak{q}} R/\mathfrak{q}[x_1, \ldots, x_d] \to \text{gr}_{\mathfrak{q}} M,$$

where $R/\mathfrak{q}[x_1, \ldots, x_d]$ is the polynomial ring on d variables over $R/\mathfrak{q}$, making $x_1, \ldots, x_d$ act like a set of generators for $\mathfrak{q}$. Then show that any nonzero graded submodule of $M/\mathfrak{q}M \otimes_{R/\mathfrak{q}} R/\mathfrak{q}[x_1, \ldots, x_r]$ has dimension r.

Exercise 12.12a: Do induction on the number of generators of $\mathfrak{q}$, using the exact sequence

$$0 \to (0 :_A x_1)(-1) \to A(-1) \to A \to A/(x_1) \to 0.$$

Chapter 13

Exercise 13.2: The coefficients of the characteristic polynomial of b are the elementary symmetric functions in the conjugates of b.

Exercise 13.3: Let g_i be a set of generators of T as an algebra over k. First form the ring S generated by the elementary symmetric functions in the conjugates of the g_i, and show that T is integral over it. Thus T, and hence also T^G, is a finitely generated S-module.

Exercise 13.7: Exercise 2.15c shows that P is prime. To show that $\text{codim}\, Q/P = 1$, we may first factor out P and invert all the nonzero homogeneous elements in the resulting ring. Thus we may assume that every homogeneous element is a unit. Now S_0 is a field, and $S \cong S_0$ or $S_0[x, x^{-1}]$ by Exercise 2.18.

Exercise 13.8:

1a. Given an ideal J of R, set $J' = JR[x, x^{-1}] \cap \mathcal{R}_I(R)$. Show that $(J_1 \cap J_2)' = J_1' \cap J_2', (I^n)' = (I')^n$; that if P is a prime ideal then P' is a prime ideal; and that if I is P-primary then I' is P'-primary. Deduce that $()'$ preserves the primary decomposition of 0.

1b. Show that $\mathcal{R}(R)/J' \cong \mathcal{R}_{(I+J)/J}(R/J)$. Use this with part 1a to reduce to the case where R is a domain. Note that the localization $\mathcal{R}_I(R)[(x^{-1})^{-1}] = R[x, x^{-1}]$; by Exercise 10.1 this shows $\dim \mathcal{R}_I(R) \geq 1 + \dim R$. The opposite inequality follows from Theorem 13.8.

2a and b. Imitate the solutions to 1a and b.

3. By Corollary 13.7 the dimension of $\mathrm{gr}_I(R)$ is the maximum of the codimensions of homogeneous prime ideals. Show that the maximal homogeneous prime ideals are the ideals of the form $P^* := P/I \oplus I/I^2 \oplus \cdots$ for P a maximal ideal of R containing I. Further, $\mathrm{gr}_I(R)_{P^*} = \mathrm{gr}_I(R_P)_{P^*}$. This reduces the problem to the case where R is local with maximal ideal $P \supset I$.

   Note that $\mathrm{gr}_I R = \mathcal{R}_I(R)/(x^{-1})\mathcal{R}_I(R)$, and that x^{-1} is not in any minimal prime of $\mathcal{R}_I(R)$. Thus $\dim \mathrm{gr}_I R \leq \dim \mathcal{R}_I(R) - 1 = \dim R$. For the opposite inequality, show that if $P \supset P_1 \supset \cdots$ is a chain of distinct primes of R, then $P' + (x^{-1}) \supset P' \supset P_1' \supset \cdots$ is a chain of distinct primes of $\mathcal{R}_I(R)$. Thus $\operatorname{codim} P' + (x^{-1}) \geq 1 + \operatorname{codim} P$. By Corollary 10.8, $\dim \mathrm{gr}_I R \geq \operatorname{codim} Q - 1$.

Exercise 13.11: It is enough to consider the image of an open set U defined by the nonvanishing of a single element $f \in A(X)$. Let $P(t)$ be a monic polynomial with coefficients in $A(Y)$ such that $P(f) = 0$. Prove that the image of U is open by showing that $Y - \varphi(U)$ is the closed set defined by the coefficients of P.

To this end write $K(X)$ and $K(Y)$ for the fields of fractions of $A(X)$ and $A(Y)$, respectively, and let $L \supset K(Y)$ be the normal closure of the field extension $K(X) \supset K(Y)$. Let T be the integral closure of $A(X)$ (or equivalently of $A(Y)$) in L, and let $\psi : X' \to X$ be the corresponding map of varieties. First note that $\psi\psi^{-1}U = U$, so that we need only consider the case $X = X'$. Show directly that a prime of $A(X)$ not containing f contracts to a prime of $A(Y)$ not containing at least one of the coefficients of the characteristic polynomial of f. Prove the converse (a prime containing f contracts to a prime containing all the coefficients) by using Proposition 13.10.

Exercise 13.12: Reduce to the case where R is local, and apply Theorems A and 13.8.

Chapter 14

Exercise 14.1, part 3: Let V be the set of polynomials of degree d obtained by multiplying each $f_i(s_1, \ldots, s_m; x_0, \ldots, x_n)$ by all the monomials in the X_i of $\deg d - \deg f_i$, and let M be the matrix of coefficients of the

polynomials in V, with respect to the basis consisting of the monomials of degree d in the X_i. M is a $p \times q$ matrix, with p equal to the number of polynomials in V, and q equal to the number of monomials of degree d in the X_i; the coefficients of M are polynomials in the s_i. For $a \in \mathbf{A}^m$, let $M(a)$ be the matrix obtained by evaluating the entries of M at a. We have $a \in X_d$ iff the polynomials in V, evaluated at $(s_1, \ldots, s_m) = a$, fail to span the set of all monomials of degree d. This happens iff $M(a)$ has rank $< q$, that is, iff all the $q \times q$ minors of $M(a)$ vanish.

Exercise 14.3: Let $Z \subset \mathbf{P}^n \times \mathbf{P}^{n^\vee}$ be the set of pairs (x, H) such that $x \in X$, $x \in H$, and H contains the tangent space to X at x. These are all closed conditions—the last may be expressed in terms of the vanishing at x of certain minors of the Jacobian matrix associated with the generators of the ideal of X restricted to L—so Z is closed. Its image under projection to $\mathbf{P}^{n^\vee}$ is X'.

Chapter 15

Exercise 15.1: It is generated by monomials.

Exercise 15.2: If the submodule is generated by monomials $\{g_i\}$, let $M_i = I_i e_i$ be the submodule generated by all the g_j that are of the form me_i.

Exercise 15.3: The given elements are certainly in $(I : n)$. On the other hand, if $f \in (I : n)$ then $fn \in I$, so the terms of fn are multiples of some m_i. It follows from unique factorization that the terms of f are multiples of some $m_i/\mathrm{GCD}(m_i, n)$.

Exercise 15.7: $I_1 \cap I_2 = I_1 I_2$ iff a minimal generating set $\{m_i\}$ for I_1 and a minimal generating set $\{n_j\}$ for I_2 do not have any variables in common. The "if" part is easy. To prove "only if," suppose on the contrary $I_1 \cap I_2 = I_1 I_2$ but $m_i = pm_i'$ and $n_j = pn_j'$ have GCD $p \neq 1$, and that m_i' is chosen with minimal degree. Since $I_1 \cap I_2 \supset pm_i'n_j'$, we see that $pm_i'n_j'$ is a multiple of some $m_u n_v$. Deduce a contradiction from the assumed minimality of degrees.

Exercise 15.12: From Exercise 15.11 it follows that zero is not in the convex hull of the finitely many elements $m_i - n_i \in P_<$. Equivalently, there is a rational hyperplane H through the origin in $\mathbf{Q}^r$ such that a translate of H separates 0 from the $m_i - n_i$. Writing H as the set of zeros of a linear functional λ, we see that λ is either strictly positive or strictly negative on all the $m_i - n_i$. In the second case we replace λ by $-\lambda$. Since we may

multiply λ by a positive integer without changing H, we may assume that λ is integral.

Exercise 15.13: Find a rational linear functional w whose hyperplane of zeros does not meet the interior of the positive cone $P_<$. If w is nonnegative on $P_<$, use w for the first weight vector; otherwise use $-w$. Do induction on the dimension of the span of $P_<$, considering the intersection of $P_<$ with the hyperplane of zeros of w.

Exercise 15.16: The sequence $\mathrm{in}(f_t')$ is nonincreasing and thus must eventually stabilize; let m_1 be its eventual value. Similarly, the sequence $\mathrm{in}(f_t' - m_1)$ must eventually stabilize; let m_2 be its eventual value, and so on. Show that if the process did not terminate, then $m_1, m_2, \ldots$ would be an infinite strictly descending sequence.

Exercise 15.18: If not all the syzygies were linear combinations of the given syzygies, we could choose one, say $\sum p_u \varepsilon_u$, with the property that the largest monomial m among the $\mathrm{in}(p_u g_u)$ is minimal. Let $\sum' p_v g_v$ be the sum of all those terms $p_v g_v$ for which $\mathrm{in}(p_v g_v)$ is m up to a scalar. Writing $\mathrm{in}(p_v g_v) = n_v \,\mathrm{in}(g_v)$ for some term n_v of p_v, we have $\sum' n_v \,\mathrm{in}(g_v) = 0$, so there is a linear combination of the given syzygies that has the form $\sum' n_v \varepsilon_v - \sum f_u \varepsilon_u$ for some f_u with $\mathrm{in}(f_u g_u) < m$. Subtract this from the syzygy $\sum p_u \varepsilon_u$ to get a contradiction.

Exercise 15.20: In this case we have $m_{ij} = \mathrm{in}(g_i)/e_i$ and $m_{ji} = \mathrm{in}(g_j)/e_j$. We have

$$\begin{aligned} -m_{ji} g_i + m_{ij} g_j &= (g_j g_i - g_i g_j) - (\mathrm{in}(g_j) g_i - \mathrm{in}(g_i) g_j) \\ &= (g_j - \mathrm{in}(g_j)) g_i - (g_i - \mathrm{in}(g_i)) g_j \\ &= p_j g_i - p_i g_j \end{aligned}$$

with $\mathrm{in}(p_j) < \mathrm{in}(g_j)$ and $\mathrm{in}(p_i) < \mathrm{in}(g_i)$. We claim that the initial term of such an expression is necessarily $\mathrm{in}(p_j)\,\mathrm{in}(g_i)$ or $\mathrm{in}(p_i)\,\mathrm{in}(g_j)$. Indeed, the only other possibility is that these terms cancel. But since $\mathrm{in}(g_i)$ and $\mathrm{in}(g_j)$ are relatively prime, cancellation is only possible if $\mathrm{in}(g_i)$ divides $\mathrm{in}(p_i)$, which is impossible because of the inequality. Now subtract the appropriate multiple of g_i or g_j, and repeat the argument.

Exercise 15.21: You need go no further than $u = 1$ and the case of two variables.

Exercise 15.24: First, if $b' \in \mathcal{B}'$ and $u \in \mathcal{U}$, observe that the upper-left $s \times s$ submatrix of $b'u$ is the product of the upper-left $s \times s$ submatrix of b' and the upper-left $s \times s$ submatrix of U. In particular, it is the product of invertible matrices, so the principal minors of $b'u$ are all nonzero.

Conversely, suppose that the principal minors of a matrix g are all nonzero. It suffices to show that there is a lower triangular matrix b such that $u = bg$ is in $\mathcal{U}$; then $g = b^{-1}u$ shows that $g \in \mathcal{B}'\mathcal{U}$. But multiplying g on the left by a lower triangular matrix may be expressed as a sequence of elementary transformations, in each of which one either multiplies a row of g by a nonzero scalar or adds a row of g to a later row. Since the 1×1 principal minor of g, which is the upper-left entry g_{11}, is nonzero we may multiply the first row by g_{11}^{-1} and then subtract a multiple of the first row from each succeeding row to make g into a matrix whose first column has entries $1, 0, \ldots, 0$. The effect of this is to multiply the principal minors by $g_{11}^{-1} \neq 0$. In particular, the principal minor of order 2, which is now equal to g_{22}, is nonzero. Multiplying the second row by g_{22}^{-1} and then adding a multiple of it to each succeeding row, we may assume that the second column of g has entries $g_{12}, 1, 0, \ldots, 0$. Continuing in this way, we eventually reduce to an element $g \in \mathcal{U}$ as claimed.

Since each of the principal minors is a polynomial function of the entries of g, the locus where they are all nonzero is open; as $\mathcal{G}$ is itself an open subset of an affine space, any open subset is dense.

Exercise 15.25: Modify the last paragraph of the given proof to use only the elements g_i from I.

Exercise 15.26a: Take $r = 2$ and $>$ as the lexicographic order. Let K be the ideal generated by all x_1/x_2^s for $s \geq 0$. Show that $x_2K = K$. Since T is a domain, Corollary 4.7 shows that K cannot be finitely generated.

Exercise 15.27: To get a Gröbner basis adjoin $g_4 = yz^2$. The syzygies on the original three generators are generated by the columns of

$$\begin{pmatrix} y^2 & 0 & (x+z)y \\ -x^2 & x+z & 0 \\ 0 & -y & -x^2 \end{pmatrix}.$$

Exercise 15.28: The resolution will be symmetric, with ranks of free modules 1, 5, 5, 1, and the first and last matrices should be transposes of one another up to change of basis; if you make the change of basis necessary to make the first and last matrices actually be transposes of one another, the middle matrix will be skew-symmetric. This phenomenon will be "explained" in Chapter 21.

Exercise 15.29: $x^2, txy + y^3, xy^3, y^5$

Exercise 15.36: Begin by homogenizing the elements of a presentation matrix for M with respect to a new variable x_0 and multiplying each element by whatever power of x_0 is necessary to bring them all to the same

degree, to get a homogeneous submodule $M'' \subset F$ whose cokernel is a graded $S[x_0]$-module. Then argue as in the proposition.

Chapter 16

Exercise 16.1: From

$$d(b) = d(b \cdot b) = b\, d(b) + b\, d(b)$$

we get $b\, db = (1-b)\, db$. The equation $bm = (1-b)m$ implies that $m = 0$ for any idempotent b and S-module M. Alternate proof: Localize and reduce to the case $b = 1$ or $b = 0$, treated in the text.

Exercise 16.2: $\Omega_{S/R} = M$.

Exercise 16.13: Take $I = (x_1^2, \ldots, x_c^2)$.

Exercise 16.14a: Show that the transcendence degree of the quotient field $\mathbf{Q}((x_1, \ldots, x_r))$ over $\mathbf{Q}$ is uncountable, and use Proposition 16.9.

Exercise 16.17a: To show that a map is an injection, it is enough to show this locally at each associated prime of 0 in the source. Since I/I^2 is free over $T = S/I$, and I is a radical ideal, these associated primes are just the minimal primes of T. Localizing at such a prime P, T becomes the field $K(T/P)$, and thus it is enough to treat the case where S is local and T is the residue class field of the regular local ring S.

Now, in general, if $U \to V \to W \to 0$ is a right-exact sequence of vector spaces, then the map $U \to W$ is a monomorphism iff $\dim V = \dim U + \dim W$, so it suffices to check this equality in the case at hand.

By Corollary 16.21 (see the remark immediately after it), $\Omega_{S/k}$ is free of rank equal to the transcendence degree of S over k. By Corollary 16.17, $\Omega_{T/k}$ is a vector space of dimension equal to the transcendence degree of T over k. The vector space I/I^2 has dimension $= \dim S$.

Writing S as $(k[x_1, \ldots, x_r]/Q)_p$, where $T = K(k[x_1, \ldots, x_r]/P$, we see that the transcendence degree of S over k is the dimension of $k[x_1, \ldots, x_r]/Q$, which is the dimension of S plus the transcendence degree of T over k, as required.

Chapter 17

Exercise 17.2: First reduce to the case where R is local, and the condition is that a, b is a regular sequence. Suppose $ax - b$ is prime. Note first that $R[x]$ is a free R-module, so that $H^1(K(a,b)) = 0$ iff $H^1(R[x] \otimes K(a,b)) = 0$, so that it is enough to show that the Koszul complex of a, b in the ring $R[x]$

is exact. But over $R[x]$ the sequence a, b can be transformed by an invertible 2×2 matrix into the sequence $ax+b$, a; if $ax+b$ is prime, this last is a regular sequence, so $H^1(K(ax+b, a)) = 0$. Now use the preceding exercise.

Exercise 17.6 (From an unpublished note of J. Sally):

Step 1: Show that if $x_1, \ldots, x_r$ is a regular sequence and x_r is a nonzerodivisor modulo $x_1, \ldots, x_i$ for each $i = 0, \ldots, r-1$, then $x_r, x_1, \ldots, x_{r-1}$ is a regular sequence. (To do this you might factor out x_1 and use induction, treating the $r = 2$ case separately.)

Step 2: Show that if $x_1, \ldots, x_{r-1}$ is a regular sequence in any order and x_r is a nonzerodivisor modulo every subset of $\{x_1, \ldots, x_{r-1}\}$ then $x_1, \ldots, x_r$ is a regular sequence in any order.

Step 3: Use prime avoidance in the form given in Exercise 3.8b to show that if $x_1, \ldots, x_r$ is a regular sequence, then there is an element $j \in (x_1, \ldots, x_{r-1})$ such that $x_r + j$ is a nonzerodivisor modulo every subset of $\{x_1, \ldots, x_{r-1}\}$.

Step 4: Complete the proof of the assertion of Exercise 17.6 by induction on r.

Exercise 17.9: Use the short exact sequences

$$0 \to \wedge^j T_1 \to \wedge^{j+1} R^{n+1}(j+1) \to \wedge^{j+1} T_1 \to 0 \quad (j = 1, \ldots, n)$$

and the long exact sequence in Koszul homology to which they give rise.

Exercise 17.11:

a. Send e_I to $(\mathrm{LCM}(m, m_1, \ldots, m_t)/\mathrm{LCM}(m_1, \ldots, m_t))e_I$, where LCM denotes the least common multiple.

b. The basis element e_I of $T(n_1, \ldots, n_t)$ corresponds to $e_{I \cup \{t+1\}}$ of $T(m_1, \ldots, m_t, m)$.

c. Do induction on t; use the long exact sequence associated to the mapping cone, and note that multiplication by m takes $R/(n_1, \ldots, n_t) = H_0(T((n_1, \ldots, n_t))$ monomorphically to $R/(m_1, \ldots, m_t) = H_0(T((m_1, \ldots, m_t))$.

Exercise 17.12a: We may map a free A/I-module onto I/I^2 sending the ith generator to x_i. To prove that this map is an isomorphism, it suffices to prove this locally, so we may suppose that A is a local ring. If $\sum_{i=1}^{c} a_i x_i = \sum_{1 \le i \le j \le c} b_{ij} x_i x_j$, we must show that all the a_i are in $(x_1, \ldots, x_c)$. Since A is local, every permutation of $x_1, \ldots, x_c$ is a regular sequence, so it suffices to show that $a_c \in I$.

To this end we rewrite the equality above as

$$(a_c - b_{cc}x_c)x_c = \sum_{1\le i\le j\le c-1} b_{ij}x_ix_j - \sum_{1\le i\le c-1} b_{ic}x_ix_c \in (x_1,\dots,x_{c-1}).$$

Since $x_1,\dots,x_c$ is a regular sequence, we have

$$a_c - b_{cc}x_c \in (x_1,\dots,x_{c-1}),$$

whence $a_c \in (x_1,\dots,x_c)$.

Exercise 17.13:

a. Use the facts that associated primes of monomial ideals are generated by subsets of the variables, and that the ideal quotient of two monomomial ideals is a monomial ideal.

b. Note that $R/I = S/J \otimes R$, and that if $J = (m_1,\dots,m_t)$, then $\mathrm{Tor}_i^S(S/J,R) = H_i(T(m_1,\dots,m_t)\otimes R)$.

c. Use the short exact sequence

$$0 \to S/(J\cap J') \to S/J \oplus S/J' \to S/(J+J') \to 0,$$

and the vanishing of $\mathrm{Tor}_1^S(S/(J\cap J'),R)$.

d. If $J' = (m_1,\dots,m_t)$, then $(J:J') = \cap_i(J:m_i)$, so using part e the problem reduces to the case where J' is principal. But $(J:m)$ is the ideal generated by the coefficients of m in syzygies $\sum a_im_i + bm = 0$; so the result follows from part d.

Exercise 17.14:

a. Let $S_j(I)$ be the jth symmetric power of I as an R-module (see Appendix for the definition). Show that $S_j(I) \cong I^j$ by induction on j, by computing a free presentation of each.

b. Use part a, inverting the element x_1 in I. (This is a strengthening of the fact that the blowup of I in R is covered by open affines like Spec $R[x_2/x_1,\dots,x_r/x_1]$).

Exercise 17.15: Prove the proposition by induction on r. First prove that if $\mathrm{in}(x)$ is a nonzerodivisor on $\mathrm{gr}_I(R)$, then x is a nonzerodivisor. Next show that if $\mathrm{in}(x)$ is a nonzerodivisor of degree d on $\mathrm{gr}_I(R)$, then $(x)\cap I^{n+d} = xI^n$ for every integer $n \ge 0$. Deduce that if $\bar I$ is the image of I in $\bar R := R/(x)$, then $\mathrm{gr}_I(R)/\mathrm{in}(x)\,\mathrm{gr}_I(R) = \mathrm{gr}_{\bar I}\bar R$.

The hypothesis that R is local is used in the first step; it would be enough to know that $\cap_n I^nR/(x_1,\dots,x_s) = 0$ for every $s \ge 0$.

Exercise 17.16:

a. First suppose that $x_1, \ldots, x_r$ is a regular sequence. It is necessary to show that I^j/I^{j+1} is isomorphic to the symmetric power $S_j(I/I^2)$, a free R/I-module with free generators the monomials in the x_i. Do this first for $(y_1, \ldots, y_r) \subset S := \mathbf{Z}[y_1, \ldots, y_r]$. Prove as in Exercise 17.13c that
$$\mathrm{Tor}_1^S(S/(y_1, \ldots, y_r)^j, R) = 0$$
and deduce the general case by tensoring the exact sequence
$$0 \to (y_1, \ldots, y_r)^j/(y_1, \ldots, y_r)^{j+1} \to S/I^{j+1} \to S/I^j \to 0$$
with S.

b. For the converse, use Exercise 17.15 with $I = (x_1, \ldots, x_r)$.

Exercise 17.17c: In the first case the homology is
$$\begin{aligned} H^0(K(x)) &= k \cdot x_1^2 x_2^2 1 \\ H^1(K(x)) &= k \cdot x_2^2 m_1 \oplus k \cdot x_1^2 m_2 \\ H^2(K(x)) &= k \cdot m_1 \wedge m_2; \end{aligned}$$

$$\begin{aligned} H_2(K'(x)) &= k \cdot x_1^2 x_2^2 n_1 \wedge n_2 \\ H_1(K'(x)) &= k \cdot x_1^2 n_1 \oplus k \cdot x_2^2 n_2 \\ H_0(K'(x)) &= k \cdot 1. \end{aligned}$$

In this case the module structure on the first is trivial—that is, $M = \wedge^1 M$ annihilates $H^*(K(x))$—while the algebra structure on the second is isomorphic to an exterior algebra on a two-dimensional vector space over k.

In the second case, if we write $\mathfrak{m}$ for the ideal (x_1, x_2), the homology is
$$\begin{aligned} H^0(K(x)) &= m^2 1 \cong kx^2 \oplus kxy \oplus ky^2 \\ H^1(K(x)) &= k \cdot x_2^2 m_1 \oplus k \cdot x_1^2 m_2 \\ H^2(K(x)) &= k \cdot m_1 \wedge m_2; \\ H_2(K'(x)) &= m^2 n_1 \wedge n_2 \cong kx^2 n_1 \wedge n_2 \oplus kxy n_1 \wedge n_2 \oplus ky^2 n_1 \wedge n_2 \\ H_1(K'(x)) &= k \cdot x_1^2 n_1 \oplus k \cdot x_2^2 n_2 \\ H_0(K'(x)) &= k \cdot 1. \end{aligned}$$

In this case the module structure on the first is nontrivial (for example, $m_1 \cdot H^0 K(x) = k \cdot x_2^2 m_1$), while the algebra structure on the second is trivial—that is, $(H_1(K'(\varphi)) \oplus H_2(K'(\varphi)))^2 = 0$.

Exercise 17.18:

a. If the linear series V is base-point free, then the complex in question is locally the Koszul complex of a sequence of elements generating

the unit ideal. Since the cohomology of $K(x_1, \ldots, x_r)$ is annihilated by $x_1, \ldots, x_r$, this proves exactness.

b. Apply part b to the sheaf $F = \mathcal{L}^{\otimes n}$ for $n \geq 2$.

Chapter 18

Exercise 18.7: For the second part, show that the elements of $k[x_0, \ldots, x_3]/I$ can each be written in the form (scalar) $\cdot x_0^a \cdot x_1^b \cdot x_2^c \cdot x_3^d$ with $0 \leq b$, $c \leq 1$. Then consider the natural map of $k[x_0, \ldots, x_3]/I$ onto R.

Exercise 18.8: It helps to think of this as a subring of $k[s^4, s^3t, s^2t^2, st^3, t^4]$, which is Cohen-Macaulay.

Exercise 18.11: If

$$r_0x^n + r_1x^{n-1} + \cdots + r_n = (s_0x^d + s_1x^{d-1} + \cdots)(t_0x^e + t_1x^{e-1} + \cdots),$$

with $s_0, t_0 \neq 0$, then because R is a domain we must have $d + e = n$. It follows that $s_0, t_0 \notin P$, and, by a descending induction, that $s_i, t_j \in P$ for $i, j \geq 1$. But then $r_n = s_dt_e \in P^2$, contradicting our hypothesis.

Exercise 18.18: First show that it is enough to prove that R is flat over the localization $k[x_1, \ldots, x_r]_{(x_1, \ldots, x_r)}$. You can then use the local criterion of flatness.

Chapter 19

Exercise 19.3: If R has finite dimension, show that R is regular iff R has finite global dimension. Reduce to the case where R is local (and thus of finite dimension). If M is an R-module such that $R[x] \otimes M$ is projective, show that M is projective; deduce that if $R[x]$ is regular then M is. For the converse, note that if R is local then graded modules over $R[x, y]$ have minimal free resolutions like modules over a local ring. If R is regular local, show first that graded modules over $R[x, y]$ have finite free resolutions, and deduce that arbitrary modules over $R[x]$ have finite free resolutions.

Exercise 19.4: Suppose that $a_i \in I_i$ and $b_i \in J_i$ are nonzero divisors, and let

$$c = \prod a_i \prod b_i.$$

We first show that $m = n$. If we invert c, then each I_i and J_i becomes isomorphic to R, and we are reduced to showing that over a commutative ring, the rank of a free module is well defined; this may be deduced from the fact that $\wedge^i R^n = 0$ iff $i \geq n$ (or see Corollary 4.4).

To show that the products are equal, we shall show that $I_1 \cdot \cdots \cdot I_n$ is isomorphic to the image of $\wedge^n(I_1 \oplus \cdots \oplus I_n)$ in $\wedge^n(I_1 \oplus \cdots \oplus I_n) \otimes R[c^{-1}]$ under the localization map. Since the isomorphism of $I_1 \oplus \cdots \oplus I_n$ with $J_1 \oplus \cdots \oplus J_n$ induces an isomorphism of $\wedge^n(I_1 \oplus \cdots \oplus I_n)$ with $\wedge^n(J_1 \oplus \cdots \oplus J_n)$, this will have the desired consequence.

To accomplish this, we embed each of the ideals I_i in R, so that we get a map $I_1 \oplus \cdots \oplus I_n \to R^n$ which becomes an isomorphism upon inverting c. Taking nth exterior powers, we get a map $\wedge^n(I_1 \oplus \cdots \oplus I_n) \to \wedge^n R^n = R$. By the first statement of Proposition 7.4, $\wedge(I_1 \oplus \cdots \oplus I_n) = \wedge I_1 \otimes \cdots \otimes \wedge I_n$, so that in particular $\wedge^n(I_1 \oplus \cdots \oplus I_n)$ is a direct sum of terms of which one is $\wedge^1 I_1 \otimes \cdots \otimes \wedge^1 I_n = I_1 \otimes \cdots \otimes I_n$ and the rest each involve some exterior power $\wedge^k I_i$ with $k > 1$. The map

$$\wedge^1 I_1 \otimes \cdots \otimes \wedge^1 I_n \to \wedge^1 R \otimes \cdots \otimes \wedge^1 R = \wedge^n R^n = R$$

induced by the inclusion is just the product map, so that its image is $I_1 \cdot \cdots \cdot I_n$. On the other hand, if $k > 1$, then

$$\wedge^k I_i \otimes R[c^{-1}] = \wedge^k R \otimes R[c^{-1}] = 0,$$

so the image of $\wedge^n(I_1 \oplus \cdots \oplus I_n)$ in $\wedge^n(I_1 \oplus \cdots \oplus I_n) \otimes R[c^{-1}]$ is the same as the image of $I_1 \otimes \cdots \otimes I_n$ in $\wedge^n(I_1 \oplus \cdots \oplus I_n) \otimes R[c^{-1}] = R[c^{-1}]$ under the product map—that is, since c is a non zerodivisor, $I_1 \cdot \cdots \cdot I_n$, and we are done with the second statement.

Exercise 19.6:

a. Consider an arbitrary nonzero map $P \to R$. Its image is an ideal I. Since I is projective, we see that $P = I \oplus P'$, and P' has smaller rank than P.

c. First do the case where R is local, and thus a DVR; the proof is then the same as for the fundamental theorem of abelian groups. Note that this includes the case of a module over a factor ring of a DVR. Then set $J = \operatorname{ann} M$. Show that J is a nonzero ideal. Since R is Dedekind, R/J is Artinian, and thus a product of its localizations, which are factor rings of DVRs. Split M according to this product decomposition, and reduce M to the desired form separately over the factors. Put the results together using Exercise 2.27.

Exercise 19.12: Locally at (p) we have $M_{(p)} = \mathbf{Z}_{(p)} \cdot (1/p) \subset \mathbf{Q}$. Show that the composite map $M \to M_{(p)} = \mathbf{Z}_{(p)} \cdot (1/p) \cong \mathbf{Z}_{(p)} \to \mathbf{Z}/(p)$ does not lift to a map $\varphi : M \to \mathbf{Z}$ by considering $1 \in \mathbf{Z} \subset M$ and showing that $\varphi(1)$ would have to be divisible by every prime.

Exercise 19.18: If M has a free resolution

$$0 \to F_m \to \cdots \to F_1 \to F_0 \to M \to 0,$$

with $F_i = \oplus S(-a_{ij})$, then we must have

$$c_t(M) = \prod_{i\,\text{even},\,j} (1 - a_{ij}t) / \prod_{i\,\text{odd},\,j} (1 - a_{ij}t) \bmod t^{r+1}.$$

Chapter 20

Exercise 20.3:

a. The usual structure theorem for modules over a principal ideal domain shows that any module is uniquely a direct sum of cyclic modules whose annihilators form a chain of ideals. Compute the Fitting ideals in terms of a presentation adapted to this structure.

b. One module has an element annihilated by z and the other doesn't.

Exercise 20.5: Show that any one minor of φ is a linear combination of the appropriate minors of φ' by restricting to a ring finitely generated over $\mathbf{Z}$, and to finitely generated free modules mapping to F and F'.

Exercise 20.7: First do the case where the elements are the variables of a polynomial ring. In this case use the fact that the only ideals of $k[x_1, \ldots, x_n]$ that are invariant under the natural action of GL_n are the powers of the maximal ideal. Show that because of the form of the answer, it determines the ideals $I\varphi_k$ in the general case.

Exercise 20.11: What closed sets are defined by the $\mathrm{Fitt}_i M$?

Exercise 20.12: Show that for each k, $\{P \text{ a prime} \mid \operatorname{rank} M_P = k\}$ is open.

Exercise 20.13: By Lemma 19.2 it is enough to treat the case where R, P is local. Show that $\mu_P(M) = \mu_Q(M)$ for every minimal prime Q of R. Let $n = \mu_P(M)$. By Nakayama's lemma M can be generated by n elements, that is, there is a surjection $\varphi : R^n \to M$. Show that φ becomes a monomorphism on tensoring with $K(R/Q)$ for each minimal prime Q. Use the fact that R is reduced to conclude that φ is an isomorphism. For part b, consider the R-module R_{red}.

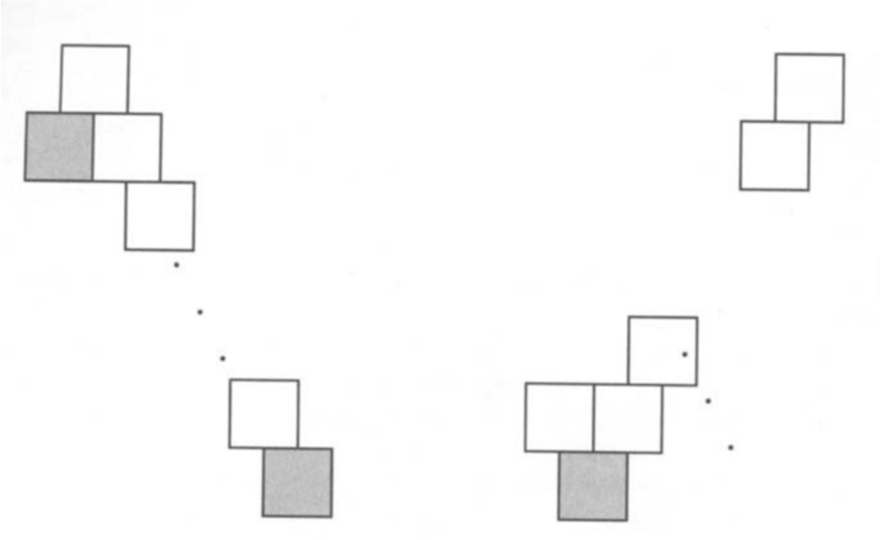
Exercise 20.20: It follows from local duality that if $S = k[x_0, \ldots, x_n]$ is the homogeneous coordinate ring of $\mathbf{P}^n$, then M has depth ≥ 2 (so $\mathrm{Ext}^{n+1-j}(M, S) = 0$ for $0 \leq j \leq 2$) and

$$H^{j-1}(\mathbf{P}^n; \mathcal{F}(m-j))^{\check{}} \cong \mathrm{Ext}^{r+1-j}(M, S)_{-r+j-m}$$

for $j \geq 2$. The given formula is now a direct translation of Proposition 20.16.

Chapter 21

Exercise 21.1: Here are the pictures for A and $D(A)$, with the socles shaded:



Exercise 21.2:

1. Show that D takes each simple module to itself (use $\operatorname{ann} M \subset \operatorname{ann} D(M) \subset \operatorname{ann} D^2(M) = \operatorname{ann} M$).

2. By induction on the length of a module M, show that length M = length $D(M)$.

3. If $M \to N$ is an epimorphism and $D(N) \to D(M)$ is not a monomorphism, then $D(N) \to D(M)$ factors through a module of smaller length. Apply D again and derive a contradiction. Similarly, show that D takes monomorphisms to epimorphisms.

4. Finish the problem by proving: If the complex $0 \to M' \to M \to M'' \to 0$ is exact at M' and at M'' and if length M = length M' + length M'', then the complex is exact.

Exercise 21.4: The functor $D = \operatorname{Ext}^d(-, R)$ is certainly R-linear and contravariant. If $0 \to M' \to M \to M'' \to 0$ is a short exact sequence of modules of finite length, then $\operatorname{Ext}^{d-1}(M', R) = 0$ by Theorem 18.4 while $\operatorname{Ext}_R^{d+1}(M', R) = 0$ by Corollary 19.6. Thus from the long exact sequence in Ext we get a short exact sequence

$$0 \to \operatorname{Ext}_R^d(M', R) \to \operatorname{Ext}_R^d(M, R) \to \operatorname{Ext}_R^d(M, R) \to 0,$$

proving that D is exact.

It thus remains to show that there are natural isomorphisms $M \to \operatorname{Ext}_R^d(\operatorname{Ext}_R^d(M, R), R)$ for all R-modules of finite length M. Let

$$\mathcal{F} : 0 \to F_c \to F_{c-1} \to \cdots \to F_1 \to F_0$$

be a free resolution of M. Because M has finite length, the annihilator of M has depth d, so the modules $\operatorname{Ext}_R^i(M, R) = 0$ for all $i < d$ by Proposition 18.4. This implies that the complex $\mathcal{F}^* = \operatorname{Hom}_R(\mathcal{F}, R)$, whose homology is $\operatorname{Ext}_R(M, R)$, is a free resolution of $\operatorname{Ext}_R^d(M, R)$. Repeating this argument, we see that $\mathcal{F}^{**}$ is a free resolution of $\operatorname{Ext}_R^d(\operatorname{Ext}_R^d(M, R), R)$. Since all the free modules in $\mathcal{F}$ may be chosen to be finitely generated, the natural homomorphism $\mathcal{F} \to \mathcal{F}^{**}$ is an isomorphism, and induces the required natural isomorphism $M \to \operatorname{Ext}_R^d(\operatorname{Ext}_R^d(M, R), R)$. (The reader may check that this isomorphism is independent of the resolutions chosen.) □

Exercise 21.5: By what we have already done, the dualizing functor is unique on the category of modules over a fixed zero-dimensional factor ring of R, and thus on the category of modules of any fixed length. Show that the isomorphism between two isomorphic dualizing functors on one of these subcategories is already determined by the isomorphism on the simple module.

Exercise 21.10: From the definitions, if $a, b \in A$ and $w \in \omega_A$, then $\psi(aw)(b) = \eta(abw) = (a\eta)(bw)$, so $\psi(aw) = a\psi(w)$, and ψ is a homomorphism of A-modules.

Applying the duality functor $\operatorname{Hom}_k(-, k)$ we see that the dual of the map $\psi_\eta : \omega_A \to \operatorname{Hom}_k(A, k)$ is the map

$$A = \operatorname{Hom}_k(\operatorname{Hom}_k(A, k), k) \to \operatorname{Hom}_k(\omega_A, k)$$

sending 1 to η. Since $\omega_A \cong \operatorname{Hom}_k(A, k)$, its dual $\operatorname{Hom}_k(\omega_A, k)$ is isomorphic to A. Since ψ_η is an isomorphism iff its dual is, this proves the equivalence of statements 1 and 2.

For the equivalence of statements 2 and 3, let $\alpha : A/P \hookrightarrow \omega_A$ be the inclusion of the socle in ω_A. The dual of α is a projection $A \cong \operatorname{Hom}_k(\omega_A, k) \twoheadrightarrow \operatorname{Hom}_k(A/P, k) \cong A/P$. This projection carries $\eta \in \operatorname{Hom}_k(\omega_A, k)$ to $\eta\alpha \in \operatorname{Hom}_k(A/P, k)$. Thus the projection is nonzero on η iff η is nonzero on the socle of ω_A. But up to multiplication by a unit, there is only one projection $A \twoheadrightarrow A/P$, and the elements carried to nonzero elements of A/P are precisely the units of A. These are the elements that generate A as an A-module. Thus η is nonzero on the socle of ω_A iff the image of η generates ω_A.

Exercise 21.16: First check the case $c = 0$ by induction on r; then do induction on c.

Exercise 21.18:

a. Suppose first that $\dim A = 0$. In this case ω_A is isomorphic to an ideal iff $\omega_A \cong A$, because ω_A and A have the same length.

   In general, suppose first that ω_A is isomorphic to an ideal. Localizing and using the zero-dimensional case, we see that A is generically Gorenstein.

   If A is generically Gorenstein, show that the first fitting ideal of ω_A contains a nonzerodivisor u, and that after inverting u every localization of A is Gorenstein. Conclude that $\omega_A[u^{-1}]$ is an invertible module over $A[u^{-1}]$. Construct a homomorphism $\varphi : \omega_A \to A$ that is an injection after inverting u, and conclude that φ is an injection by showing that u must be a nonzerodivisor on ω_A.

b. Show that u may be chosen to be homogeneous, and proceed as in part a.

Exercise 21.19: The condition on the Hilbert series implies that $h_{\omega_A}(t) = (-1)^{\dim A} t^n h_A(t^{-1}) = t^m h_A(t)$ for some m. We wish to prove that ω_A is generated by a single element.

Use Exercise 21.18b to conclude that for some integer $e, \omega_A(e)$ is isomorphic, as a graded module, to a homogeneous ideal I.

Deduce from the hypothesis on $h_{\omega_A}(t)$ that $\dim_k(\omega_A)_m = 1$. Thus $\dim_k I_{m-e} = 1$. Choose a nonzero element $f \in I_{m-e}$.

We have $Af \subset I$. Show that the Hilbert series of Af is $t^{m-e}h_A(t)$, and that this is the same as the Hilbert series of I. Conclude that $I = Af$.

Exercise 21.20: Suppose that I is a canonical module. I has codimension ≥ 1 since the annihilator of I is 0. I has codimension ≤ 1 and A/I is Cohen-Macaulay since $\operatorname{depth}(P, A/I) \geq \operatorname{depth}(P, I) - 1$ by Corollary 18.6. To show that A/I is Gorenstein, compute $\omega_{A/I} = \operatorname{Ext}^1_A(A/I, \omega_A) = \operatorname{Ext}^1_A(A/I, I)$. From the exact sequence $0 \to I \to A \to A/I \to 0$ we get a long exact sequence including

$$\begin{aligned} \operatorname{Hom}_A(A/I, I) \to \operatorname{Hom}_A(A, I) &\to \operatorname{Hom}_A(I, I) \\ &\to \operatorname{Ext}^1_A(A/I, I) \to \operatorname{Ext}^1_A(A, I), \end{aligned}$$

that is,

$$0 \to I \to A \to \omega_{A/I} \to 0,$$

so $\omega_{A/I} = A/I$.

Exercise 21.21: Let I be an ideal of A isomorphic to the canonical module. I is unmixed of codimension 1, so I is principal (generated by the GCD of its elements).

Appendix 1

Exercise A1.1: Show that it suffices to do the case where R is a field and S is a finitely generated reduced k-algebra. Embed S in its total quotient ring, which is a finite direct product of fields. Show that tensor products commute with finite direct products. Finally, apply Theorem A1.3.

Exercise A1.2:

a. It suffices to treat the case where R and S are finitely generated fields over k. By the characterization of Theorem A1.3 b), if $k \subset k'$ is any extension of fields such that $R \otimes_k k'$ is a domain, then the quotient field of $R \otimes_k k'$ is separable over k'. For this reason it suffices to treat the case where $S = k(z)$ is generated by one element over k.

   If z is transcendental over k, then the result is obvious. If z is algebraic over k, satisfying an irreducible polynomial $p(Z) = 0$, say, then to show that $R \otimes_k S = R[Z]/(p(Z))$ is a field it suffices to show that p is irreducible over R. But the factors of p have coefficients that are algebraic over k; since k is algebraically closed in R, any factors in $R[Z]$ would be in $k[Z]$.

b. Reduce to the following special case: If S is a field of characteristic p and $s \in S$, then $S[x]/(x^p - s)$ is a local ring.

Appendix 2

Exercise A2.3a: Use Nakayama's lemma and Proposition A2.2.

Exercise A2.9: The symmetric algebras of R-modules are the quotients $R[x_1, \ldots,]/I$ where I is generated by elements homogeneous of degree 1 in the x_i.

Exercise A2.11c: A "quick and dirty" argument: First consider the case where $\alpha = e_1 \wedge e_2 + e_3 \wedge e_4 + \cdots$. Reduce to this case by first reducing to the case of a generic alternating matrix over $\mathbf{Z}[x_{ij}]$, then embedding this ring into $\mathbf{C}$ so that the problem is reduced to the case of a "generic" matrix of complex numbers, and finally reducing this complex matrix to the form φ_α with α as above.

Exercise A2.13: Use Nakayama's lemma and base change to compute the numbers of generators of the exterior and symmetric powers of I.

Though accessible to bare-handed computation, this exercise can be done more easily with a little extra technique:

(1) The presentation given is part of the Tate resolution of the residue class field.

(2) R is a one-dimensional Gorenstein ring, so any torsion-free R-module is reflexive.

(3) Because the multiplicity of R is 2, and R is one-dimensional, every ideal of R can be generated by (at most) two elements (they are free modules of rank 2 over a polynomial ring in one variable over which R is integral).

Exercise A2.16: Suppose φ is a matrix for which $\mathcal{C}^i(\varphi)$ is exact, for some $i < -1$. Deduce from Theorem A2.10, Exercise 10.10, and the criterion of exactness (Theorem 20.9) that $\mathcal{C}^i$ would be exact in the generic case and would also be a resolution of a module whose annihilator has codim $i(f - g + i)$, the codimension of the generic determinantal ideal. Derive a contradiction from Proposition 18.4.

Appendix 3

Exercise A3.2:

a. First show that the image of any map from a finitely generated module to an infinte direct sum is contained in a finite subsum. If R is Noetherian, prove using Lemma A3.4 that the direct sum of injectives is injective. For the converse, suppose that $I_1 \subset I_2 \subset \cdots$ is an ascending chain of ideals with union I. Embed each I/I_j into an injective module E_j, and extend the map $I \to \oplus E_j$ to R.

b. Choose a maximal direct sum of indecomposable injectives contained in a given injective. Use primary decomposition to produce another indecomposable summand.

Exercise A3.4a: To show that E' is injective, it suffices to show that φ can be extended to a map $R \to E'$. The kernel of φ certainly contains PI. By the Artin-Rees lemma, $P^d \cap I \subset PI$ for some d. Thus φ factors through $I \to I/(I \cap P^d) = P^d + I/P^d$. Thus we may extend φ to a map $R/P^d \to E$, and use this to define an extension $R \to R/P^d \to E$ of φ to R. The image of this extension is contained in the image of R/P^d, which is annihilated by P^d and thus contained in E'.

b: Set $Q := \mathrm{Hom}_{\mathrm{gr}}(R, k)$ and let $P = \oplus_{d>0} R_d$ be the homogeneous maximal ideal of R. Write $Q^{(d)} = \mathrm{Hom}_{\mathrm{gr}}(R/P^d, k)$. We have $Q^{(d)} = \mathrm{Hom}_k(R/P^d, k)$ because R/P^d is a finite dimensional vector space over k, so $Q^{(d)}$ is injective over R/P^d. Furthermore, $Q = \cup_d Q^{(d)}$. If I is any ideal, and $\alpha : I \to Q$ is any homomorphism, then since I is finitely generated the image of I is contained in some $Q^{(d)}$. For this value of d the map α factors through a map $\bar{\alpha} : I/P^d \cap I \to Q^{(d)}$. The map $\bar{\alpha}$ can be extended to a map

$R/P^d \to Q^{(d)}$, and composing this with the projection $R \to R/P^d$, we get a map that extends α. Thus Q is injective. To show that it is the injective hull of $k = Q^{(0)}$, show directly from the definition that k is an essential submodule.

Exercise A3.6b: First do the case where $0 \subset M$ is not the intersection of two nonzero submodules. Reduce to this case by Noetherian induction, as in the proof of the existence of primary decomposition.

Exercise A3.7a: One direction is Lemma A3.8. For the other direction, show that if $F \subset E$ is essential, then so is $\mathrm{Hom}_R(S, F) \subset \mathrm{Hom}_R(S, E)$.

Exercise A3.10: If we regard the columns of the diagram as complexes, then the first of these results is a special case of the long exact sequence in homology coming from a short exact sequence of complexes.

Exercise A3.13: It suffices by symmetry to prove the first isomorphism. If we write $K = \ker(F \oplus G \to M)$, then the inclusion of F into $F \oplus G$ induces an inclusion of exact sequences

$$\begin{array}{ccccccccc} 0 & \to & N_F & \to & F & \to & M & \to & 0 \\ & & \downarrow & & \downarrow & & \| & & \\ 0 & \to & K & \to & F \oplus G & \to & M & \to & 0. \end{array}$$

The snake lemma now shows that $G \cong \mathrm{coker}(N_F \to K)$. Since G is projective, we obtain $K \cong N_F \oplus G$ as required.

Exercise A3.15: Do induction on i, and use an exact sequence

$$0 \to A' \to A \to A'' \to 0,$$

with A projective.

Exercise A3.16: Consider the resolution

$$0 \to R \xrightarrow{x} R$$

of R/x.

Exercise A3.22a: Reduce to the case where there is a "surjective comparison map" between the two presentations—that is, a commutative diagram

$$\begin{array}{ccccccc} F & \to & G & \to & M & \to & 0 \\ \downarrow & & \downarrow & & \| & & \\ F' & \to & G' & \to & M & \to & 0 \end{array}$$

where the downward maps are surjective.

Exercise A3.25: Let E^0 be an injective module containing R/I, and set $T = E/(R/I)$. Note first that $\mathrm{Hom}_R(I, R/I) = \mathrm{Hom}_R(I/I^2, R/I)$. If

$$E : E^0 \to E^1 \to \cdots$$

is an injective resolution of R/I, show that maps from R/I into T are the same a (-1)-cycles of the complex $\mathrm{Hom}_R(R/I, E)$. Show that via the short exact sequence $0 \to I \to R \to R/I \to 0$, maps of R/I to T, mod those factoring through E_0, correspond to maps $I \to R/I$. Alternately, the relation $\mathrm{Ext}^1_R(R/I, R/I) = \mathrm{Hom}_{R/I}(\mathrm{Tor}_1^R(R/I, R/I), R/I)$ follows from the "change-of-rings" spectral sequence $\mathrm{Ext}^*_{R/I}(\mathrm{Tor}^R_*(R/I, R/I), R/I) \Rightarrow \mathrm{Ext}^*_R(R/I, R/I)$ and the fact that $\mathrm{Tor}_0^R(R/I, R/I) = R/I$ is a free R/I-module. This same spectral sequence shows that *all* the $\mathrm{Ext}^i_R(R/I, R/I)$ are dual to the $\mathrm{Tor}_i^R(R/I, R/I)$ if the latter are all free over R/I —for example, if R is local and I is its maximal ideal, or if I is generated by a regular sequence.

Exercise A3.28: The multiplication on P^* corresponds to a comultiplication $\Delta : P \to P \otimes P$. The condition on d^* guarantees the commutativity of the diagrams

$$\begin{array}{ccccc}
P_{m+n+1} & \stackrel{\Delta}{\to} & \Sigma_{i+j=m+n+1} P_i \otimes P_j & \stackrel{\text{project}}{\longrightarrow} & P_{m+1} \otimes P_n \\
\downarrow & & & & \downarrow \\
P_{m+n} & \stackrel{\Delta}{\to} & \Sigma_{i+j=m+n} P_i \otimes P_j & \stackrel{\text{project}}{\longrightarrow} & P_m \otimes P_n.
\end{array}$$

The product of cocycles $f : P_m \to k$ and $g : P_n \to k$ is the composite

$$g \circ f = h : P_{m+n} \stackrel{\Delta}{\to} \Sigma_{i+j=m+n} P_i \otimes P_j \stackrel{\text{project}}{\longrightarrow} P_n \otimes P_m \stackrel{g \otimes f}{\longrightarrow} k \otimes k = k.$$

From the commutativity of the diagram above, it follows that h is again a cocycle.

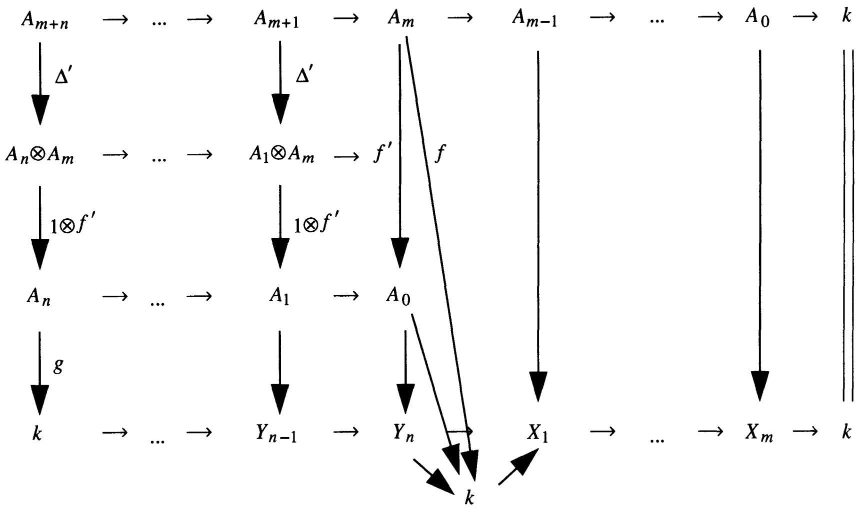
Now with notation as in Exercise A3.27, suppose that f represents the class of α and g represents the class of β. To show that $g \circ f$ represents the class of $\beta\alpha$, it suffices to check the commutativity of the following diagram. In this diagram we have written Δ' for the composition of Δ and the projection onto one component of $A \otimes A$, and f' for the lifting of f to A_0. Its commutativity follows easily from the commutativity of the smaller diagram given at the beginning of this hint.

Exercise A3.29a: By Exercise A3.22c it is enough to show that the induced map $N \otimes A \to N \otimes B$ is a monomorphism for every finitely generated module N. Reduce by localizing and factoring out a power of the maximal ideal to the statement: If $A \to B \to C \to 0$ is a right-exact sequence of modules of finite length, and $B \cong A \oplus C$, then $A \to B$ is a monomorphism. This follows at once by comparing

$$\mathrm{length}\ker(B \twoheadrightarrow C) = \mathrm{length}\, B - \mathrm{length}\, C$$

to length A.

Exercise A3.44a: To define G_0, choose generators m_i^* for $\mathrm{gr}_{\mathfrak{m}} M$ and lift them back to a (possibly nonminimal) set of generators m_i for M, so that



m_i^* is the leading form of m_i in $\mathrm{gr}_{\mathfrak{m}} M$. Let $G_0 \to M$ be the map from a free module with basis corresponding to these generators. Filter G_0 in such a way that its free generators have the same degree in $\mathrm{gr}\, G_0$ as the corresponding m_i^* have in $\mathrm{gr}_{\mathfrak{m}} M$. Filter the kernel M_1 of $G_0 \to M$ by taking the induced filtration from G_0, that is, $M^p := (G_0)^p$. This is a stable filtration by the Artin-Rees lemma. Repeat this process inductively for the rest of the resolution.

Exercise A3.46c: Use the Auslander-Buchsbaum formula (Chapter 19) and Theorem 18.4.

Exercise A3.49: Break up F into short exact sequences

$$0 \to K^i \to F^i \to B^i \to 0$$

and

$$0 \to B^i \to K^{i+1} \to H^{i+1}(F) \to 0.$$

Use the same idea as in Proposition A3.16 to form a short exact sequence of injective resolutions for the modules of each of these short exact sequences, and put these together.

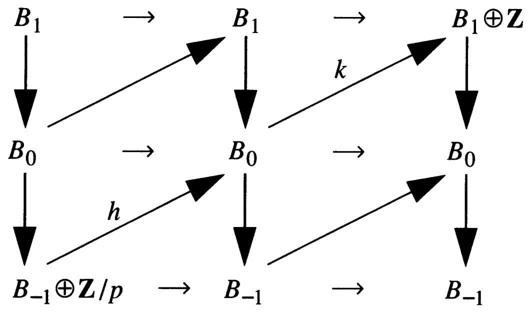
Exercise A3.50:

b. Use the following double complex: Take an acyclic resolution of A, apply $\mathcal{F}$, and take an injective resolution of the resulting complex in the sense of Exercise A3.49.

d. Use resolutions by flabby sheaves.

Exercise A3.51: If $\alpha : A \twoheadrightarrow B$ and $\beta : B \rightarrowtail C$ were such a factorization, and if $\gamma : B' \to B$ is such that $B' \to B \to C$ is homotopic to 0, then γ

would have to be homotopic to 0 itself. Apply this to the mapping cone of α, and apply the dual argument to the mapping cone of β, to get a diagram such as the following,



where the left and right columns are (parts of) the mapping cones of α and β, the horizontal maps give maps of complexes that are homotopic to 0, and the diagonal maps are the homotopies. Deduce that the maps labelled h and k give a surjection $\mathbf{Z}/p \to \mathbf{Z}$, a contradiction.

Appendix 6

Exercise A6.10: Consider the Snake Lemma (Appendix 3).

References

Akin, K, D.A. Buchsbaum, and J. Weyman (1982). Schur functors and Schur complexes. *Adv. in Math.* 44, pp. 207–278.

Akizuki, Y. (1935). Einige Bemerkungen über primäre Integritätsbereiche mit Teilerkettensatz. *Proc. Phys.-Math. Soc. Japan* 17, pp. 327–336.

Altman, A. and S. Kleiman (1970). *Introduction to Grothendieck Duality Theory.* Springer Lect. Notes in Math 146, Springer-Verlag, New York.

André, M. (1974). *Homologie des algèbres commutatives.* Springer-Verlag, New York.

Andreotti, A. and T. Frankel (1959). The Lefschetz theorem on hyperplane sections. *Annals of Math.* 69, pp. 713–717.

Andreotti, A. and P. Salmon (1957). Annelli con unica decomponibilità in fattori primi ed un problema di intersezioni complete. *Monatsh. für Math.* 61, pp. 97–142.

Anick, D. (1982). Counterexample to a conjecture of Serre. *Annals of Math.* 115, pp. 1–33.

Anick, D. (1988). Recent progress in Hilbert and Poincaré series. In *Algebraic Topology: Rational Homotopy* (Louvain-la-Neuve, 1986), pp. 1–25. Springer Lect. Notes in Math. 1318, Springer-Verlag, New York.

Apéry, R. (1945a). Sur certain charactères numériques d'un idéal sans composant impropre. *C. R. Acad. Sci. Paris, Ser. A-B* 220, pp. 234–236.

Apéry, R. (1945b). Sur les courbes de première espèce de l'espace à trois dimensions. *C. R. Acad. Sci. Paris, Ser. A-B* 220, pp. 271–272.

Arbarello, E., M. Cornalba, P. Griffiths and J. Harris (1985). *Geometry of Algebraic Curves.* Springer-Verlag, New York, NY.

Artin, E. and J.T. Tate (1951). A note on finite ring extensions. *J. Math. Soc. Japan* 3, pp. 74–77.

Artin, E. (1965). *The Collected Papers of E. Artin.* Ed. S. Lang and J.T. Tate. Addison-Wesley, Reading, MA.

Artin, M. (1973). *Théoremes de représentabilité pour les espaces algébriques.* Les Presses de l'Univ. de Montréal, Montréal, Canada.

Artin, M. (1976). *Deformations of Singularities.* Tata Inst. Lect. Notes, Bombay, India.

Artin, M. (1991). *Algebra.* Prentice Hall, Englewood Cliffs, NJ.

Atiyah, M.F. and I.G. Macdonald (1969). *Introduction to Commutative Algebra.* Addison-Wesley, Reading, MA.

Auslander, M. (1955). On the dimension of modules and algebras III. *Nagoya Math. J.* 9, pp. 67–77.

Auslander, M. (1966). Coherent functors. In *Proceedings of the Conference on Categorical Algebra,* La Jolla, 1965, Ed. S. Eilenberg *et al.* Springer-Verlag, New York.

Auslander, M. and D.A. Buchsbaum (1956). Homological dimension in Noetherian rings. *Proc. Natl. Acad. Sci. U.S.A.* 42, pp. 36–38.

Auslander, M. and D.A. Buchsbaum (1958). Codimension and multiplicity. *Annals of Math.* 68, pp. 625–657.

Auslander, M. and D.A. Buchsbaum (1959). Unique factorization in regular local rings. *Proc. Natl. Acad. Sci. U.S.A.* 45, pp. 733–734.

Avramov, L.L. (1975). Flat morphisms of complete intersections. *Sov. Math. Doklady* 16, pp. 1413–1417.

Avramov, L., H.-B. Foxby and B. Herzog (1994). The structure of local homomorphisms. *J. Alg.* 164, pp. 124–145.

Avramov, L., V. Gasharov and I. Peeva (1994). *Complete intersection dimension.* Preprint.

Avramov, L. and S. Halperin (1986). Through the looking-glass: A dictionary between rational homotopy and local algebra. In *Algebraic, Algebraic Topology, and Their Interactions*, Proc. Stockholm 1983, Ed. J.-E. Roos. Springer Lect. Notes in Math. 1183, Springer-Verlag, New York.

Avramov, L. and J. Herzog (1994) Jacobian criteria for complete intersections. The graded case. *Invent. Math.* 117, pp. 75–88.

Azumaya, G. (1950). On maximally central algebras. *Nagoya Math. J.* 2, pp. 119–150.

Baer, R. (1940). Abelian groups that are direct summands of every containing abelian group. *Bull. Amer. Math. Soc.* 46, pp. 800–806.

Bayer, D. (1982). The division algorithm and the Hilbert scheme. Thesis, Harvard University, Cambridge, MA.

Bayer, D. and I. Morrison (1988). Standard bases and geometric invariant theory I: Initial ideals and state polytopes. *J. Symb. Comp.* 6, pp. 209–217.

Bayer, D. and M. Stillman (1987a). A theorem on refining division orders by the reverse lexicographic orders. *Duke J. Math.* 55, pp. 321–328.

Bayer, D. and M. Stillman (1987b). A criterion for detecting m-regularity. *Invent. Math.* 87, pp. 1–11.

Bayer, D. and M. Stillman (1988). On the complexity of computing syzygies. *J. Symb. Comp.* 6, pp. 135–147.

Bayer, D. and M. Stillman (1992). Computation of Hilbert functions. *J. Symb. Comp.* 14, pp. 31–50.

Bayer, D. and M. Stillman (1982–1990). Macaulay: A system for computation in algebraic geometry and commutative algebra. Source and object code available for Unix and Macintosh computers. Contact the authors, or download from zariski.harvard.edu via anonymous ftp. (login:anonymous, password: any, cd Macaulay).

Bass, H. (1963). On the ubiquity of Gorenstein rings. *Math. Z.* 82, pp. 8–28.

Beauville, A. (1983). *Complex projective surfaces.* London Math. Soc. Lect. Note series, 68. Cambridge Univ. Press, Cambridge, England.

Bell, J.L. and A.B. Slomson (1969). *Models and Ultraproducts: An Introduction.* North-Holland Publ. Co., Amsterdam.

Bergman, G. (1978). The diamond lemma for ring theory. *Adv. in Math.* 29, pp. 178–218.

Bott, R. and L. Tu (1982). *Differential Forms in Algebraic Topology.* Springer-Verlag, New York.

Bourbaki, N. (1983). *Algébre Commutative*, Chapters 8–9. Masson, New York.

Bourbaki, N. (1985). *Commutative Algebra*, Chapters 1–7. Springer-Verlag, New York.

Bourbaki, N. (1970). *Algèbre I*, Chapters I–III. Hermann, Paris, France, = *Algebra I*, Chapters I–III. English translation: Springer-Verlag, 1989.

Bourbaki, N. (1981). Algèbre, Ch. IV–VII. Masson, Paris, France, = Algebra II, Ch. 4–7. English translation: Springer-Verlag, 1990.

Braun, R. and G. Fløystad. A bound for the degree of smooth surfaces in P^4 not of general type. *Compositio Math.* To appear.

Brodmann, M.P. and R.Y. Sharp (In Press). *Local cohomology: an algebraic introduction with geometric applications.* Cambridge University Press. Cambridge, U.K.

Bruns, W. (1976). "Jede" endliche freie Auflösung ist freie Auflösung eines von drei Elementen erzeugten Ideals. *J. Algebra* 39, pp. 429–439.

Bruns, W. and J. Herzog (1993). *Cohen-Macaulay Rings.* Cambridge University Press, Cambridge, U.K.

Bruns, W. and U. Vetter (1988). *Determinantal Rings.* Springer Lect. Notes. in Math. 1327. Springer-Verlag, New York.

Buchberger, B. (1970). An algorithmic criterion for the solvability of algebraic systems of equations. *Aequationes Math.* 4, pp. 374–383.

Buchberger, B. (1976). A theoretical basis for the reduction of polynomials to canonical forms. *ACM SIGSAM Bull.* 39, pp. 19–29.

Buchberger, B. (1987). History and basic features of the critical pair completion procedure. *J. Symb. Comp.* 3, pp. 3–38.

Buchsbaum, D.A. (1964). A generalized Koszul complex I. *Trans. Amer. Math. Soc.* 111, pp. 183–196.

Buchsbaum, D.A. (1970). Complexes associated with the minors of a matrix. *Symposia Math. IV* (Istituto Nazionali di Alta Matematica), pp. 255–283. Academic Press, London.

Buchsbaum, D.A. and D. Eisenbud (1973). What makes a complex exact? *J. Algebra* 25, pp. 259–268.

Buchsbaum, D.A. and D. Eisenbud (1973). Remarks on ideals and resolutions. *Symposia Math. XI*, pp. 193–204. Academic Press, London.

Buchsbaum, D.A. and D. Eisenbud (1974). Some structure theorems for finite free resolutions. *Advances in Math.* 12, pp. 84–139.

Buchsbaum, D.A. and D. Eisenbud (1977). What annihilates a module? *J. Alg.* 47, pp. 231–243.

Buchsbaum, D.A. and D. Eisenbud (1977b). Algebraic structures for finite free resolutions and some structure theorems for ideals of codimension 3. *Amer. J. Math.* 99, pp. 447–485.

Buchsbaum, D.A. and D.S. Rim (1964). A generalized Koszul complex, II. Depth and multiplicity. *Trans Amer. Math. Soc.* 111, pp. 197–224.

Buchweitz, R., D. Eisenbud, and J. Herzog (1987). Cohen-Macaulay modules on quadrics. In *Singularities, Representations of Algebras, and Vector Bundles*, pp. 58–116. Ed. G.M. Greuel and G. Trautmann. Springer Lect. Notes in Math. 1273, Springer-Verlag, New York.

Cartan, H. (1954). Exposé 7, *Sem. Cartan* 7.

Cartan, H. and S. Eilenberg (1956). *Homological Algebra.* Princeton University Press, Princeton, NJ.

Cayley A. (1848). On the theory of elimination. Cambridge and Dublin *Math. Journal Vol. III* pp. 116–120. Reprinted in Collected Math. Papers I, pp. 370–374. Cambridge University Press, Cambridge, 1889.

Chase, S.U. (1960). Direct products of modules. *Trans. Amer. Math. Soc.* 97, pp. 457–473.

Chevalley, C. (1944). On the theory of local rings. *Annals of Math.* 44, pp. 690–708.

Chevalley, C. (1958). *Variétés complètes. Fondements de la géométrie algébrique* (Chapter IV). Secr. Math. of the Institut Henri Poincaré, Paris, France.

Choi, S. (1988). The divisor class group of surfaces of embedding dimension 3. *J. Algebra* 119, pp. 162–169.

Ciliberto, C. J. Harris and R. Miranda (1988). General components of the Noether-Lefschetz locus and their density in the space of all surfaces. *Math. Ann.* 282, pp. 667–680.

Clebsch, A. (1864). Zur Theorie der algbraischen Flächen. *J. für reine und angew. Math.* 63, pp. 14–26.

Cohen, I.S. (1946). On the structure and ideal theory of complete local rings. *Trans. Amer. Math. Soc.* 59, pp. 54–106.

Cohen, I.S. and A. Seidenberg (1946). Prime ideals and integral dependence. *Bull. Amer. Math. Soc.* 52, pp. 252–261.

Cohn, P.M. (1966). Some remarks on the invariant basis property. *Topology* 5, pp. 215–228.

Conca, A., and J. Herzog. (In press) Ladder determinantal ideals have rational singularities. *Adv. in Math.*

Cook, M. The connectedness of space curve invariants. *Compos. Math.* To appear.

Cornell, G. and J. Silverman, Eds. (1986). *Arithmetic Geometry.* Springer-Verlag, New York.

Cox, D. J. Little, and D. O'Shea (1992). *Ideals, Varieties, and Algorithms.* Springer-Verlag, New York.

DeConcini, C. D. Eisenbud and C. Procesi (1980). Young diagrams and determinantal varieties. *Invent. Math.* 56, pp. 129–165.

DeConcini, C. D. Eisenbud and C. Procesi (1982). *Hodge Algebras. Asterisque 91, Soc. Math. de France.*

Dickson, L.E. (1921). Determination of all homogeneous polynomials expressible as determinants. *Trans. Amer. Math. Soc.* 22, pp. 167–179.

Dolgachev, I. and M. Kapranov (1993). Schur quadrics, cubic surfaces and rank 2 vector bundles over the projective plane. *Journées de Géométrie Algébrique d'Orsay* (Orsay, 1992). Astérisque 218 (1993), pp. 111–144.

Eakin, P. (1968). The converse to a well-known theorem on Noetherian rings. *Math. Ann.* 177, pp. 278–282.

Eagon, J. and M. Hochster (1974). *R*-sequences and indeterminates. *Quarterly J. of Math.* 25, pp. 61–71.

Eagon, J. and D.G. Northcott (1962). Ideals defined by matrices, and a certain complex associated to them. *Proc. Royal Soc.* 269, pp. 188–204.

Edwards, H.M. (1977). *Fermat's Last Theorem; A Genetic Introduction to Algebraic Number Theory.* Springer-Verlag, New York.

Eckmann, B. and A. Schopf (1953). Über injektive Moduln. *Archiv der Math.* 4, pp. 75–78.

Eisenbud, D. (1970). Subrings of Artinian and Noetherian rings. *Math. Ann.* 185, pp. 247–249.

Eisenbud, D. (1980). Homological algebra on a complete intersection. *Trans. Amer. Math. Soc.* 260, pp. 35–64.

Eisenbud, D. (1988). Linear sections of determinantal varieties. *Amer. J. Math* 110, pp. 541–575.

Eisenbud, D. (1992). Green's conjecture: An orientation for algebraists. In *Free Resolutions in Commutative Algebra and Algebraic Geometry*, Ed. D. Eisenbud and C. Huneke. Jones and Bartlett, Boston, MA.

Eisenbud, D. and S. Goto (1984). Linear free resolutions and minimal multiplicity. *J. of Algebra* 88, pp. 89–133.

Eisenbud, D. and M. Green (1994). Minors of matrices in free resolutions. *Duke J. of Math.*

Eisenbud, D. and J. Harris (1987). On varieties of minimal degree (A centennial account). In *Algebraic Geometry, Bowdoin 1985*, Ed. S. Bloch. *Amer. Math. Soc. Symp. in Pure and App. Math.* 46, pp. 1–14.

Eisenbud, D. and J. Harris (1992). *Schemes, the Language of Modern Algebraic Geometry.* Chapman-Hall, New York, NY

Eisenbud, D. and M. Hochster (1979). A Nullstellensatz with nilpotents. *J. of Algebra* 58, pp. 157–161.

Eisenbud, D., C. Huneke and W. Vasconcelos (1992). Direct methods for primary decomposition. *Invent. Math.* 110, pp. 207–235.

Eisenbud, D. and J. Koh (1991). Some linear syzygy conjectures. *Adv. in Math.* 90, pp. 47–76.

Eisenbud, D. and H. Levine (1977). An algebraic formula for the topological degree of a C^∞ map germ. *Annals of Math.* 106, pp. 19–44.

Eisenbud, D. A. Reeves and B. Totaro (1995). Initial ideals of Veronese subrings. *Adv. in Math.*

Eisenbud, D. and D. Saltman (1989). Rank varieties of matrices. In *Commutative Algebra*, pp. 173–212. Math. Sci. Res. Inst. Publ. 15, Springer, New York.

Eisenbud, D. and B. Sturmfels (1994). Binomial ideals. Preprint.

Evans, E.G. and P. Griffith (1985). *Syzygies.* London Mathematical Society Lecture Notes 106. Cambridge University Press, Cambridge, U.K.

Fauvel, J. and J. Gray (1987). *The History of Mathematics: A Reader.* Macmillan Education in association with the Open University, Basingstoke, Hampshire, U.K.

Fitting, H. (1936). Die Determinantenideale eines Moduls. *Jahresbericht der Deutschen Math.-Vereinigung* 46, pp. 195–229.

Fitchas, N. and A. Galligo (1990). Le Nullstellensatz effectif et la conjecture de Serre (theorème Quillen-Suslin) pour le calcul formel. *Math. Nachr.* 149, pp. 231–253.

Flenner, H. (1977). Die Sätze von Bertini für lokale Ringe. *Math. Ann.* 229, pp. 97–111.

Flenner, H. and W. Vogel (1993). On multiplicities of local rings. *Manuscr. Math.* 78, pp. 85–97.

Fogarty, J. (1969). *Invariant Theory.* Lect. Notes in Math. W.A. Benjamin, New York.

Formanek, E. (1973). Faithful Noetherian modules. *Proc. Amer. Math. Soc.* 41, pp. 381–383.

Freyd, P. (1964). *Abelian Categories; An Introduction to the Theory of Functors.* Harper and Row, New York.

Fulton, W. (1969). *Algebraic Curves; An Introduction to Algebraic Geometry.* W.A. Benjamin, New York.

Fulton, W. (1983). *Introduction to Intersection Theory in Algebraic Geometry.* CBMS Lect. Notes 54. Amer. Math. Soc., Providence, RI.

Fulton, W. (1984). *Intersection Theory.* Springer-Verlag, New York.

Fulton, W. (1993). *Introduction to Toric Varieties.* Annals of Math Studies 131, Princeton University Press, Princeton, NJ.

Fulton, W. and J. Harris (1991). *Representation Theory, A First Course.* Springer-Verlag, New York.

Gaeta, F. (1952). Quelques progrès récents dans la classification des variétés algébriques d'un espace projectif. In *Deuxième Colloque de Géometrie Algébrique Liège.* C.B.R.M.

Galligo, A. (1974). A propos du théorème de préparation de Weierstrass. In *Fonctions des Plusieurs Variables Complexes*, pp. 543–579. Lect. Notes in Math. 409 Springer-Verlag, New York.

Gelfand, S. and Yu.I. Manin (1989). Homological algebra. *Current Problems in Mathematics. Fundamental Directions*, 38 (Russian), pp. 5–240, Itogi Nauki i Tekhniki, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow. To appear in an English translation, Springer-Verlag, New York.

Gianni, P., B., Trager, and G. Zacharias (1989). Gröbner bases and primary decomposition of polynomial ideals. In *Computational Aspects of Commutative Algebra*, pp. 15–33. Ed. L. Robbiano, Academic Press, New York.

Gilmer, R. (1974). Dimension sequences of commutative rings. In *Ring Theory*, pp. 31–46. Ed. B.R. McDonald, A.R. Magid, and K.C. Smith. Marcel Dekker Lect. Notes in Math. 7.

Gimigliano, A. (1989). On Veronesean surfaces. *Nederl. Akad. Wetensch. Indag. Math.* 51, pp. 71–85.

Godement, R. (1958). *Topologie algébrique et théorie des faisceaux.* Hermann, Paris, France.

Gordan, P. (1900). Les invariants des formes binaires. *Journal de Mathématiques Pures et Appliqués* 6, pp. 141–156.

Gorenstein, D. (1952). An arithmetic theory of adjoint plane curves. *Trans. Amer. Math. Soc.* 72, pp. 414–436.

Goto, S. and K. Watanabe (1978). On graded rings, I. *J. Math. Soc. Japan* 30, pp. 179–213.

Grace, J.H. and A. Young (1903). *The Algebra of Invariants.* Cambridge University Press, Cambridge, U.K.

Grauert, H. (1972). Über die Deformation isolierter Singularitäten analytischer Mengen. *Invent. Math.* 15, pp. 171–198.

Green, M. (1984a). Koszul cohomology and the geometry of projective varieties I. *J. Diff. Geom.* 19, pp. 125–171.

Green, M. (1984b). Koszul cohomology and the geometry of projective varieties II. *J. Diff. Geom.* 20, pp. 279–289.

Green, M. (1989). Koszul cohomology and geometry. In *Lectures on Riemann Surfaces* (Trieste, Italy, 1987), pp. 177–200. World Science Publishing, Teaneck, NJ.

Green, M. (1989b) Restrictions of linear series to hyperplanes, and some results of Macaulay and Gotzmann. In *Algebraic Curves and Projective Geometry, Trento 1988.* Eds. E. Ballico and C. Ciliberto. Lect. Notes in Math. 1389, pp. 76–86.

Grell, H. (1927). Beziehungen zwischen den Idealen verschiedener Ringe. *Math. Ann.* 97, pp. 490–593.

Griffiths P. and J. Harris (1978). *Principles of Algebraic Geometry.* John Wiley and Sons, New York, NY.

Grivel P.-P. (1987). Les foncteurs de la catégorie des faisceaux associés à une application continu. In *Algebraic D-modules* Ed. A. Borel, et al., Academic Press, Boston, MA.

Gröbner, W. (1939). Über die algebraischen Eigenschaften der Integrale von linearen Differentialgleichungen mit konstanten Koeffizienten. *Monatsh. der Math.* 47, pp. 247–284.

Gröbner, W. (1950). Über die Eliminationstheorie. *Monatsh. der Math.* 54, pp. 71–78.

Grothendieck, A. (1957) Sur quelques point d'algèbre homologique. *Tohoku Math. J.* 9 pp. 119–221.

Grothendieck, A. (1961a). Éléments de la géométrie algébrique II, Étude globale élémentaire de quelques classes de morphismes. *Publ. Math. de l'I.H.E.S* 8.

Grothendieck, A. (1961b). Éléments de la géométrie algébrique III, Étude cohomologique des faisceaux cohérents (première partie). *Publ. Math. de l'I.H.E.S* 11.

Grothendieck, A. (1962). Techniques de construction en géométrie analytique IV. Formalisme général des foncteurs représentables. In Séminaire H. Cartan, *Familles d'espaces complexes et fondements de la géométrie analytique.* Secretariat Math. Institut Henri Poincaré, Paris, France.

Grothendieck, A. (1964). Éléments de la géométrie algébrique IV, Étude locale des schémas et des morphismes de schémas (première partie). *Publ. Math. de l'I.H.E.S* 20.

Grothendieck, A. (1965). Éléments de la géométrie algébrique IV, Étude locale des schémas et des morphismes de schémas (deuxième partie). *Publ. Math. de l'I.H.E.S* 24.

Grothendieck, A. (1967). *Local Cohomology.* Springer Lect. Notes in Math. 41, Springer-Verlag, New York.

Grothendieck, A. (1971). *Revêtements étales et Groupe Fondemental (SGA1).* Lect. Notes in Math. 224, Springer-Verlag, New York.

Grothendieck, A. (with P. Berthelot and L. Illusie) (1971). *Théorie des Intersections et Théorème de Riemann-Roch (SGA VI).* Lect. Notes in Math. 225, Springer-Verlag, New York.

Grothendieck, A. (with M. Artin and J.L. Verdier) (1972). *Théorie des Topos et Cohomologie Etale des Schémas (SGA IV).* Lect. Notes in Math. 269, Springer-Verlag, New York.

Gruson, L. and C. Peskine (1982). Courbes de l'espàce projectif: Variétés de secantes. In *Enumerative Geometry and Classical Algebraic Geometry* (Nice 1981), pp. 1–31. Progress in Math. 24, Birkhäuser, Boston, MA.

Gruson, L. R. Lazarsfeld and C. Peskine (1983). On a theorem of Castelnuovo, and the equations defining space curves. *Invent. Math.* 72, pp. 491–506.

Gulliksen, T. and G. Levin (1969). Homology of local rings. *Queen's Papers in Pure and Applied Math.* Vol. 20. Mimeographed notes, available from Queen's University, Kingston, Ontario.

Gunning, R.C. and H. Rossi (1965). *Analytic Functions of Several Complex Variables.* Prentice-Hall, Englewood Cliffs, NJ.

Harris, J. (1992). *Algebraic Geometry.* Springer-Verlag, New York.

Hartshorne, R. (1966). Connectedness of the Hilbert scheme. *Publ. Math. de l'I. H. E. S.* 29 pp. 5–48.

Hartshorne, R. (1966a). A property of A-sequences. *Bull. Soc. Math. France* 94, pp. 61–66.

Hartshorne, R. (1966b). *Residues and Duality.* Springer Lect. Notes in Math. 20, Springer-Verlag, New York.

Hartshorne, R. (1977). *Algebraic Geometry.* Springer-Verlag, New York.

Hartshorne, R. (1979). Algebraic vector bundles on projective spaces: a problem list. *Topology* 18, pp. 117–128.

Heitmann, R. (1993). Characterization of completions of unique factorization domains. *Trans. Amer. Math. Soc.* 337, pp. 379–387.

Heitmann, R. and R. Wiegand (1991). Direct sums of ideals. *Linear Algebra Appl.* 157, pp. 21–36.

Heinzer, W. L.J. Ratliff, Jr. and K. Shah (in press). Parametric decompositions of monomial ideals (I). Houston J. Math.

Hensel, K. and G. Landsberg (1902). *Theorie der algebraischen Funktionen einer Variabeln und ihre Anwendung auf algebraische Kurven und Abelsche Integrale.* B.G. Teubner, Leipzig, Germany.

Hermann, G. (1926). Der Frage der endlich vielen Schritte in der Theorie der Polynomideale. *Math. Ann.* 95, pp. 736–788.

Herzog, J. and E. Kunz (1971). Die Wertehalbgruppe eines lokalen Rings der Dimension 1. *Ber. Heidelberger Akad. Wiss.* II. Abh.

Herzog, J. E. Kunz, *et al.* (1971). *Der kanonische Modul eines Cohen-Macaulay-Rings.* Springer Lect. Notes in Math. 238, Springer-Verlag, New York.

Hilbert, D. (1890). Über die Theorie von algebraischen Formen. *Math. Ann.* 36, pp. 473–534.

Hilbert, D. (1893). Über die vollen Invariantensysteme. *Math. Ann.* 42, pp. 313–373.

Hilton, P.J. and U. Stammbach (1971). *A Course in Homological Algebra.* Springer-Verlag, New York.

Hirsch, M.W. (1976). *Differential Topology.* Springer-Verlag, New York.

Hochster, M. (1975): *Topics in the homological theory of modules over commutative rings.* CBMS Lect. Notes 24. American Math. Soc., Providence, RI.

Hochster, M. (1976). Big Cohen-Macaulay modules and algebras and the embeddability in rings of Witt vectors. *Proc. Queen's Univ. Conf. on Commutative Algebra. Queen's Papers in Pure and Applied Math.* vol. 42, pp. 106–195

Hochster, M. (1987). Intersection problems and Cohen-Macaulay modules. In *Algebraic Geometry, Bowdoin 1985.* Ed. S. Bloch. *Proc. Symp. Pure Math.* 46, part 2. Amer. Math. Soc., Providence, RI.

Hochster, M., C. Huneke and J. Sally, eds. (1989). *Commutative Algebra.* MSRI publications 15. Springer-Verlag, New York.

Hochster, M. and C. Huneke (1990). Tight closure, invariant theory, and the Briançon-Skoda theorem. *J. Amer. Math. Soc.* 3 pp. 31–116.

Hopkins C. (1939). Rings with minimal condition for left ideals. *Annals of Math.* 40, pp. 712–730.

Humphreys, J.E. (1975). *Linear Algebraic Groups.* Springer-Verlag, New York.

Huneke, C. (1982). Linkage and the Koszul homology of ideals. *Amer. J. Math.* 104, pp. 1043–1062.

Huneke, C. (1984). Numerical invariants of linkage. *Inv. Math.* 75, pp. 301–325.

Huneke, C. (1986). The dimension and components of symmetric algebras. *J. Algebra* 98, pp. 200–210.

Huneke, C. (1992). Uniform bounds in Noetherian rings. *Invent. Math.* 107, pp. 203–223.

Huneke, C. and B. Ulrich (1987). The structure of linkage. *Annals of Math.* 126, pp. 277–334.

Hurewicz, W. and H. Wallman (1941). *Dimension Theory.* Princeton University Press, Princeton, NJ.

Husemoller, D. (1975). *Fiber Bundles*, second ed. Springer-Verlag, New York.

Iversen, B. (1986) *Cohomology of Sheaves.* Springer-Verlag, New York.

Jaffe, D. (1989). On set-theoretic complete intersections in $\mathbf{P}^3$. *Math. Ann.* 285, pp. 165–176.

Jans, J. (1964). *Rings and Homology.* Holt, Rinehart and Winston, New York.

Kaplansky, I. (1958). Projective modules. *Annals of Math.* 68, pp. 372–377.

Kaplansky, I. (1970). *Commutative Rings.* Allyn and Bacon, Boston, MA. Rev. ed. The University of Chicago Press.

Kapranov, M. M., I.M. Gelfand, and A. Zelevinsky (1994). *Hyperdeterminants, Resultants, and Multidimensional Determinants.* Birkhäuser, Boston MA.

Kelly, J.L. (1955). *General Topology.* Van Nostrand, New York.

Kempf, G. (1990). Some wonderful rings in algebraic geometry. *J. Algebra*, 134, pp. 222–224.

Kirby, D. (1974). A sequence of complexes associated with a matrix. *J. London Math.* Soc. 7, pp. 523–530.

Kleiman, S.L. and A. Thorup (1994). A geometric theory of the Buchsbaum-Rim multiplicity. *J. Algebra* 167 pp. 168–231.

Kleiman, S.L. (1971). Les théorèmes de finitude pour le foncteur de Picard, Exposé XIII. In *Théorie des Intersections et Théorème de Riemann-Roch (SGA VI)*, pp. 616–666. Springer Lect. Notes in Math. 225, Springer-Verlag, New York.

Kline, M. (1972). *Mathematical Thought from Ancient to Modern Time.* Oxford University Press, Oxford, U.K.

Knuth, D. (1969). *The Art of Computer Programming. Vol. 2: Seminumerical Algorithms.* Addison-Wesley, Reading, MA.

Koh, J.H. (1988). Superheight of an ideal in a Noetherian ring. *J. Algebra* 116, pp. 1–6.

Kollár, J. (1988). Sharp effective Nullstellensatz. *J. Amer. Math. Soc.* 1, pp. 963–975.

Kraft, H.-P. (1985). *Geometrische Methoden in der Invariantentheorie.* Vieweg, Braunschweig, Germany.

Kronecker, L. (1881). Grundzüge einer arithmetischen Theorie der algebraischen Grössen. Jour. für die Reine und Angew. Math. 92, pp. 1–122.

Krull, W. (1928). Primidealketten in allgemeinen Ringbereichen. *S.-B. Heidelberg Akad. Wiss.* 7.

Krull, W. (1937). Beiträge zur Arithmetik kommutativer Integritätsbereiche III. *Math. Z.* 42, pp. 745–766.

Krull, W. (1935). *Idealtheorie.* Springer-Verlag, New York. Second edition, 1968.

Krull, W. (1938). Dimensionstheorie in Stellenringen. Jour. für die Reine und Angew Math. 179, pp. 204–226.

Kunz, E. (1974). Almost complete intersections are not Gorenstein rings. *J. Algebra* 28, pp. 111–115.

Kunz, E. (1985). *Introduction to Commutative Algebra and Algebraic Geometry.* Birkhauser, Boston, MA.

Kunz, E. (1991). Ebene algebraische Kurven. *Der Regensburger Trichter* 23. Universität Regensburg, Fakultät für Mathematik, Regensburg, Germany.

Kunz, E. (1992). Über den n-dimensionalen Residuensatz. Jahresber. *Deutsch. Math.-Verein.* 94, pp. 170–188.

Kunz, E. and R. Waldi (1988). Regular differential forms. *Contemporary Mathematics* 79. American Mathematical Society, Providence, RI.

Lam, T.Y. (1978). *Serre's Conjecture.* Springer Lect. Notes in Math. 635, Springer-Verlag, New York.

Lang, S. (1993). *Algebra*, third ed. Addison-Wesley, Reading, MA.

Larfeldt, T. and C. Lech (1981). Analytic ramifications and flat couples of local rings. *Acta Math.* 146, pp. 201–208.

Lascoux, A. (1978). Syzygies des variétés déterminantales. *Adv. in Math.* 30, pp. 202–237.

Lasker, E. (1905). Zur Theorie der Moduln und Ideale. *Math. Ann.* 60, pp. 20–116.

Lê Dung Trang and B. Teissier (1981). Variétés polaires locales et classes de Chern des variétés singulières. *Ann. of Math.* 114, pp. 457–491.

Leedham-Greene, C. (1972). The class group of Dedekind domains. *Trans. Amer. Math. Soc.* 163 pp. 493–500.

Lenstra, H.W. Jr. (1979) Euclidean number fields. *I. Math. Intelligencer* 2, pp. 6–15.

Leray, J. (1946). Structure de l'anneau d'homologie d'une représentation. *C. R. Acad. Sci. Paris* 222, pp. 1419–1422.

Leray, J. (1950). L'anneau spectral et l'anneau filtré d'un espace localement compacte et d'une application continue. *J. Math. Pures et Appl.* 29, pp. 1–139.

Lipman, J. (1975). Unique factorization in complete local rings. In *Algebraic Geometry*, pp. 531–546. Ed. S. Bloch. Proc. Sympos. Pure Math., vol. 29, Amer. Math. Soc., Providence, RI.

Lipman, J. (1984). Dualizing sheaves, differentials and residues on algebraic varieties. *Asterisque* 117, Soc. Math. de France.

Lipman, J. (1995). Notes on Derived Categories and Derived Functors, Preprint.

Löfwal, C. (1986). On the subalgebra generated by the one-dimensional elements in the Yoneda Ext-algebra. In *Algebra, Algebraic Topology, and their Interactions.* Ed. J.-E. Roos. Springer Lect. Notes in Math. 1183, Springer-Verlag, New York.

Logar, A. and B. Sturmfels (1992). Algorithms for the Quillen-Suslin theorem. *J. Algebra* 145, pp. 231–239.

Lyndon, R.C. (1946). The cohomology theory of group extensions. Harvard Univ. Thesis.

Lyndon, R.C. (1948). The cohomology theory of group extensions. *Duke Math. J.* 15, pp. 271–292.

Macaulay, F.S. (1913). On the resolution of a given modular system into primary systems including some properties of Hilbert numbers. *Math. Ann.* 74, pp. 66–121.

Macaulay, F.S. (1916). *Algebraic Theory of Modular Systems.* Cambridge Tracts 16, Cambridge University Press, Cambridge, U.K.

Macaulay, F.S. (1927). Some properties of enumeration in the theory of modular systems. *Proc. London Math. Soc.* 26, pp. 531–555.

Mac Lane, S. (1939). Modular fields (I). *Duke Math. J.* 5, pp. 372–393.

Mac Lane, S. (1963). *Homology.* Springer-Verlag, New York.

Mac Lane, S. (1971). *Category Theory for the Working Mathematician.* Springer-Verlag, New York.

MacRae, R. (1965). On an application of Fitting invariants. *J. Algebra* 2, pp. 153–169.

Manin, Y.I. (1988). *Quantum Groups and Non-Commutative Geometry.* Centre de Recherches Math., Univ. de Montréal, Montréal, Quebec.

Martin-Deschamps, M. and D. Perrin (1990). Sur la classification des courbes gauches. *Astérisque* 184–185.

Massey, W.S. (1952). Exact couples in algebraic topology. *Annals of Math.* 56, pp. 363–396.

Matsumura, H. (1970). *Commutative Algebra.* W. A. Benjamin, New York.

Matsumura, H. (1986). *Commutative Ring Theory.* Cambridge Studies in Adv. Math. 8, Cambridge University Press, Cambridge, U.K.

McAdam, S. (1974). Finite coverings by ideals. In *Ring Theory*, Proceedings of the Oklahoma Conference, pp. 163–171. Ed. B.R. McDonald, A.R. Magid, and K.C. Smith. Marcel Dekker Lect. Notes in Math. 7.

Milne, J.S. (1980). *Étale Cohomology.* Princeton University Press, Princeton, NJ.

Milnor, J. (1971). *Introduction to Algebraic K-Theory.* Princeton University Press, Princeton, NJ.

Milnor, J. and J. Moore (1965). On the structure of Hopf algebras. *Annals of Math.* 81, pp. 211–264.

Miyata, T. (1967). Note on direct summands of modules. *J. Math. Kyoto University* 7, pp. 65–69.

Möller, H.M. and F. Mora (1984). *Upper and Lower Bounds for the Degree of Standard Bases.* Ed. John Fitch. Springer Lect. Notes in Computer Science 174.

Morin, B. (1975). Calcul Jacobien. *Ann. Sci. École Norm. Sup.* 8, pp. 1–98.

Mumford, D. (1966). *Lectures on Curves on an Algebraic Surface.* Annals of Math. Studies 59, Princeton University Press, Princeton, NJ.

Mumford, D. (1975). *Curves and their Jacobians.* University of Michigan Press, Ann Arbor, MI.

Mumford, D. (1976). *Complex Projective Varieties.* Springer-Verlag, New York.

Mumford, D. and J. Fogarty (1982). *Geometric Invariant Theory*, second ed. Springer-Verlag, New York.

Nagata, M. (1962). *Local Rings.* Wiley, New York.

Newton, I. (1737). A treatise on the method of fluxions and infinite series, with its application to the geometry of curved lines. London. Reprinted in *The mathematical works of Isaac Newton I.* Ed. D. T. Whiteside. Johnson Reprint Corp., New York, 1964.

Nielsen, H.A. (1981). Free resolutions of tensor forms. In *Young Tableaux and Schur Functions in Algebra and Geometry* (Torun, 1980), Astérisque 87–88, pp. 289–302 Soc. Math. France, Paris.

Noether, E. (1921). Idealtheorie in Ringbereichen. *Math. Ann.* 83, pp. 24–66.

Noether, E. (1923). Eliminationstheorie und allgemeine Idealtheorie. *Math. Ann.* 90, pp. 223–261.

Noether, E. (1926). Der Endlichkeitsatz der Invarianten endlicher linearer Gruppen der Charakteristik p. *Nachr. Ges. Wiss. Göttingen*, pp. 28–35.

Noether, E. (1927). Abstrakter Aufbau der Idealtheorie in algebraischen Zahl- und Funktionenkörpern. *Math Ann.* 96, pp. 26–61.

Northcott, D.G. (1976). *Finite Free Resolutions.* Cambridge Tracts in Math. 71. Cambridge University Press, Cambridge, U.K.

Pardue, K. (1994). *Nonstandard Borel-Fixed Ideals.* Brandeis Ph.D. Thesis.

Peskine, C. (in press) *An Algébraic Introduction to Complex Projective Geometry 1. Commutative Algebra.* Cambridge University Press, Cambridge, U.K.

Peskine, C. and L. Szpiro (1974). Liaison des variétés algébriques I. *Invent. Math.* 26, pp. 271–302.

Poincaré, H. (1899). Complément á l'analysis situs. *Rendiconti di Palermo* 13, pp. 285–343.

Pragacz, P. and J. Weyman (1985). Complexes associated with trace and evaluation. Another approach to Lascoux's resolution. *Adv. in Math.* 57, pp. 163–207.

Priddy, S.B. (1970). Koszul resolutions. *Trans. Am. Math. Soc.* 152, pp. 39–60.

Puiseux, V.-A. (1850). Recherches sur les fonctions algébriques. *Jour. de Math.* 15, pp. 365–480.

Purkert, W. and H.J. Ilgauds (1985). *Georg Cantor.* B.G. Teubner, Leipzig, Germany.

Quillen, D. (1970). On the (co-)homology of commutative rings. *Proc. Symp. Pure Math.* 17, pp. 65–87.

Rabinowitch, S. (1929). Zum Hilbertschen Nullstellensatz. *Math. Ann.* 102, p. 520.

Rao, A.P. (1979). Liaison among curves in $\mathbf{P}^3$. *Invent. Math.* 50, pp. 205–217.

Rees, D. (1956). Two classical theorems of ideal theory. *Proc. Camb. Phil. Soc.* 52, pp. 252–253.

Rees, D. (1957). The grade of an ideal or module. *Proc. Camb. Phil. Soc.* 53, pp. 28–42.

Reid, M. (1988) *Undergraduate Algebraic Geometry.* Cambridge University Press, Cambridge, U.K.

Room, T.G. (1938). *The Geometry of Determinantal Loci.* Cambridge University Press, Cambridge, U.K.

Robbiano, L. (1986). On the theory of graded structures. *J. Symb. Comp.* 2, pp. 139–170.

Robbiano, L. and M. Sweedler (1990). Subalgebra bases. In *Commutative Algebra*, pp. 61–87. Ed. W. Bruns and A. Simis. Springer Lect. Notes in Math. 1430.

Rohn, K. and L. Berzolari (1921–1928). Algebraische Raumkurven und abwickelbare Flächen. In *Enzyklopädie der Mathematischen Wissenschaften*, Band 3 Teil 2, zweite Hälfte/Teilband A, pp. 1229–1426. B.B. Teubner, Leipzig, Germany.

Rotman, J. (1979). *An Introduction to Homological Algebra.* Academic Press, Cambridge, MA.

Rückert, W. (1932). Zum Eliminationsproblem der Potenzreihenideale. *Math. Ann.* 107, pp. 259–281.

Saint-Donat, B. (1973). On Petri's analysis of the linear system of quadrics through a canonical curve. *Math. Ann.* 206, pp. 157–175.

Saltman, D. (1982). Generic Galois extensions and problems in field theory. *Adv. in Math.* 43, pp. 250–283.

Samuel, P. (1951). La notion de multiplicité en algèbre et en géométrie algébrique. *J. Math. Pures Appl.* 30, pp. 159–274.

Scheja, G. and U. Storch (1972). Differentielle Eigenschaften der Lokalisierungen analytischer Algebren. *Math. Ann.* 197, pp. 137–170.

Schlessinger, M. (1968). Functors of Artin rings. *Trans. Amer. Math. Soc.* 130, pp. 208–222.

Schreyer, F.-O. (1980). *Die Berechnung von Syzygien mit dem verallgemeinerten Weierstrass'schen Divisionssatz.* Diplom Thesis, University of Hamburg, Germany.

Schreyer, F.-O. (1986). Syzygies of canonical curves and special linear series. *Math. Ann.* 275, pp. 105–137.

Schreyer, F.-O. (1991). A standard basis approach to syzygies of canonical curves. *J. reine angew. Math.* 421, pp. 83–123.

Seidenberg, A. (1974). Constructions in algebra. *Trans. Amer. Math. Soc.* 197, pp. 273–313.

Seidenberg, A. (1984). On the Lasker-Noether decomposition theorem. *Amer. J. Math.* 106, pp. 611–638.

Sernesi, E. (1986). *Topics on Families of Projective Schemes.* Queen's Papers in Pure and Applied Mathematics, 73. Queen's University, Kingston, Ontario, Canada.

Serre, J-P. (1955–1956). Géométrie algébrique et géométrie analytique. *Ann. Inst. Fourier.* Grenoble 6, pp. 1–42.

Serre, J-P. (1957). *Algèbre Locale. Multiplicités.* Springer Lect. Notes in Math. 11. Springer-Verlag, New York.

Serre, J-P. (1973). *Cohomologie Galoisienne.* Lect. Notes in Math. 5, Springer-Verlag New York.

Serre, J-P. (1979). *Local Fields.* Trans. M.J. Greenberg. Springer-Verlag, New York.

Shafarevich, I. R. (1972). *Basic Algebraic Geometry.* Springer-Verlag, New York.

Shelah, S. (1974). Infinite Abelian groups, Whitehead problem and some constructions. *Israel J. Math.* 18, pp. 243–256.

Silverman, J. (1986). *The Arithmetic of Elliptic Curves.* Springer-Verlag, New York.

Spear, D.A. (1977). A constructive approach to commutative ring theory. *Proc. of the 1977 MACSYMA Users' Conf.* NASA CP-2012, pp. 369–376.

Stanley, R. (1978). Hilbert functions of graded algebras. *Adv. in Math.* 28, pp. 57–83.

Stanley, R. (1979). Invariants of finite groups and their applications to combinatorics. *Bull. Amer. Math. Soc.* (N.S.) 1, pp. 475–511.

R. Steinberg: (1994) Nagata's example. In preparation

Stückrad, J. and W. Vogel (1986). *Buchsbaum Rings and Applications.* VEB Deutscher Verlag der Wissenschaften, Berlin, Germany.

Sturmfels, B. (1993). *Algorithms in Invariant Theory.* Springer-Verlag, New York.

Sturmfels, B., N.V. Trung and W. Vogel (In Press). Bounds on the degrees of projective schemes. *Math. Annalen.*

Sylvester, J.J. (1904–12). *The Collected Mathematical Papers of James Joseph Sylvester.* Cambridge University Press, Cambridge, England.

Szpiro, L. (1985). Présentation de la théorie d'Arakelov. Current trends in arithmetical algebraic geometry (Arcata, CA), *Contemp. Math.*, 67 (1987) pp. 279–293, Amer. Math. Soc., Providence, RI.

Taylor, D. (1966). Ideals generated by monomials in an R-sequence. Thesis. Chicago University.

Teichmüller, O. (1936). p-Algebren. *Deutsche Math.* 1, pp. 362–388.

Teissier, B. (1988). Monomes, volumes et multiplicités. In *Introduction à la Théorie des Singularités*, II, pp. 127–141. Travaux en Cours, 37, Hermann, Paris.

Teissier, B. (1990). Resultats recents d'algèbre commutative effective. Seminaire Bourbaki, 1989/90. Asterisque Vol. 189–190, pp. 107–131.

Thie, P. R. (1967). The Lelong number of a point of a complex analytic set. *Math. Ann.* 172, pp. 269–312.

Towber, J. (1979). Young symmetry, the flag manifold, and representations of $\mathrm{GL}_n\mathbf{Q}$. *J. Algebra* 61, pp. 414–462.

Uzkov, A.I. (1948). On rings of quotients of commutative rings. *Mat. Sbornik* (N.S.) 13, pp. 71–78.

Valla, G. (1979). Remarks on generalized analytic independence. *Math. Proc. Camb. Phil. Soc.* 85, pp. 281–289.

Van der Waerden, B.L. (1971). *Moderne Algebra*, eighth ed. Springer-Verlag, New York.

Vasconcelos, W. (1991). Computing the integral closure of an affine domain. *Proc. Amer. Math. Soc.* 113, pp. 633–638.

Vasconcelos, W. (Preprint, 1993). *Computational Methods in Commutative Algebra and Algebraic Geometry.*

Vasconcelos, W. (1994). *Arithmetic of Blowup Algebras.* Cambridge University Press, Cambridge, U.K.

Verdier, J.-L. (1977). Derived categories. Thesis. Reprinted in SGA 4 1/2, Springer Lect. Notes in Math. 569.

Vogel, W. (1984). *Results on Bezout's Theorem.* Tata Lect. Notes. Springer-Verlag, New York.

Walker, R.J. (1970). *Algebraic Curves.* Springer-Verlag, New York.

Weil, A. (1946) Foundations of Algebraic Geometry. American Math. Soc., Providence, RI.

Weyman, J. (1989). On the equations of conjugacy classes of nilpotent matrices. *Inv. Math.* 98, pp. 229–245.

White, N. ed. (1986a). *Theory of Matroids. Encyclopedia of Mathematics and Its Applications*, vol. 26. Cambridge University Press, Cambridge, U.K.

White, N. ed. (1986b). *Combinatorial Geometries. Encyclopedia of Mathematics and Its Applications*, vol. 29. Cambridge University Press, Cambridge, U.K.

White, N. ed. (1986c). *Matroid Applications. Encyclopedia of Mathematics and Its Applications*, vol. 40. Cambridge University Press, Cambridge, U.K.

Winkler, F. (1984). On the complexity of the Gröbner basis algorithm over $k[x, y, z]$. *Springer Lect. Notes in Computer Science* 174, ed. John Fitch.

Yuzvinsky, S. (1994). On generators of the module of logarithmic 1-forms with poles along an arrangement. (preprint)

Zariski, O. (1940). Algebraic varieties over ground fields of characteristic 0. *Amer. J. Math.* 62, pp. 187–221.

Zariski, O. (1943). Foundations of a general theory of birational correspondences. *Trans. Amer. Math. Soc.* 53, pp. 490–502.

Zariski, O. (1947). The concept of a simple point of an abstract algebraic variety. *Trans. Amer. Math. Soc.* 62, pp. 1–52.

Zariski, O. (1958). Introduction to the problem of minimal models in the theory of algebraic surfaces. *Publ. Math. Soc. Japan* 4, pp. 1–89.

(All the above works of Zariski are reprinted in *Oscar Zariski: Collected Papers*. Ed. M. Artin, H. Hironaka, and D. Mumford. The MIT Press, Cambridge, MA 1972.)

Zariski, O. and P. Samuel (1979). *Commutative Algebra*. Vols 1 and 2. Reprint of the 1958–60 edition. Springer-Verlag, New York.

Index of Notation

Entries are listed in order of appearance.

Index

Graduate Texts in Mathematics

continued from page ii

65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.
66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 2nd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAYEV. Probability.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.
119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.

121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.